REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* 10/Aug/2001	2. REPORT TYPE THESIS	3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
APPLYING THE METRICS THERMOSTAT TO NAVAL ACQUISITIONS FOR IMPROVING THE TOTAL OWNERSHIP COST-EFFECTIVENESS OF NEW SYSTEMS

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
2D LT MOFFITT JEFFRY D

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

8. PERFORMING ORGANIZATION REPORT NUMBER
CI01-196

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

13. SUPPLEMENTARY NOTES

14. ABSTRACT

20010904 043

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		121	19b. TELEPHONE NUMBER *(Include area code)*

Applying the Metrics Thermostat to Naval Acquisitions for Improving the Total Ownership Cost – Effectiveness of New Systems

by

Jeffry D. Moffitt, II

B.S., Operations Research
United States Air Force Academy, 1999

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN OPERATIONS RESEARCH
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2001

© 2001 Massachusetts Institute of Technology

Signature of Author: _____
Sloan School of Management
May 18, 2001

Certified by: _____
John R. Hauser
Kirin Professor of Marketing, Sloan School of Management
Thesis Supervisor

Accepted by: _____
Cynthia Barnhart
Professor of Civil and Environmental Engineering
Co-Director, Operations Research Center

- This page intentionally left blank -

Applying the Metrics Thermostat to Naval Acquisitions for Improving the Total Ownership Cost – Effectiveness of New Systems

by

Jeffry D. Moffitt, II

Submitted to the Sloan School of Management
on May 18, 2001 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Operations Research

ABSTRACT
In recent years, defense spending cuts have created a two-fold challenge for defense acquisitions organizations. First, the acquisition process must become increasingly streamlined so that overhead is minimized. Second, the acquisition process must proactively control the total ownership cost (TOC) of new systems from their conception. This preliminary research endeavors to show that by tying contractor incentives to metrics that correlate to total ownership cost drivers, DoD can manage the acquisition process with metrics, and thereby, reduce government oversight while increasing control over TOC.

This is accomplished by applying the Metrics Thermostat (MT) theory to defense acquisitions. The MT seeks to align the best interests of defense contractors with reducing TOC by prescribing a weighted set of contract incentives based on metrics that correlate to cost drivers. The MT determines a metric's weight (or incentive emphasis) according to the extent that incremental improvements in the metric can be shown to affect TOC savings. A metric's ability to affect cost savings is estimated with a hierarchy of linear regressions.

This research compiled operating and support (O&S) cost data with various cost driver metrics for 45 US Navy shipboard systems as far back as FY 1986. Preliminary results suggest that system manpower and training requirements should receive the greatest amount of emphasis in contract incentives, while system corrective maintenance is close behind. A regression including these two metrics accounted for approximately 68% of the variance in O&S cost. Underneath these two metrics in the hierarchy of cost driver metrics, the number of technical assist visit requests per system (a measure of maintainability and reliability), the natural logarithm of MTBF (a measure of reliability), the degree to which a system is automated (as assessed by those who maintain and support it), and the "sailor proofness" of a system (as assessed by those who maintain and support it) were found to exhibit highly significant relationships to manpower and corrective maintenance metrics. If structured properly, incentives based on these and other metrics, could result in substantial life cycle cost savings for the Navy and reduced procurement costs from less government oversight.

Thesis Supervisor: John R. Hauser
Title: Kirin Professor of Marketing, Sloan School of Management

Acknowledgements

Table of Contents

Figures and Tables

Chapter 1:   Introduction

1-1   Motivation:

To say that metrics are merely important to the success of an organization would be a gross understatement of the truth. More accurately, organizations *are* what they measure, and therefore, organizational success begins with choosing the right metrics. The following passage illustrates this principle well:

> Every metric, whether it is used explicitly to influence behavior, to evaluate future strategies, or simply to take stock, will affect actions and decisions. The link is simple. If a firm measures a, b, and c, but not x, y, and z, then managers begin to pay more attention to a, b, and c. Soon those managers who do well on a, b, and c are promoted or are given more responsibilities. Increased pay and bonuses follow. Recognizing these rewards, managers start asking their employees to make decisions and take actions that improve the metrics. (Often they don't even need to ask!). Soon the entire organization is focused on ways to improve the metrics. The firm gains core strengths in producing a, b, and c. *The firm becomes what it measures*" (Hauser and Katz 1998).

Since a firm actually *is* what it measures, metrics and techniques to exploit them are essential to the management of commercial firms. Recognizing this, many companies have spent large sums of money (sometimes over $100 million) on strategic initiatives that they have implemented and encouraged with metrics (Hauser 2000).

At the same time, many government organizations, particularly the Department of Defense (DoD) and the US Navy, have been struggling to find and exploit the right metrics to help them define and achieve their organizational goals. The very existence of positions such as "Command Metrics Coordinator" at Naval Sea Systems Command (NAVSEA) attests to the Navy's endeavors to employ metrics successfully.

While metrics are important to commercial firms and government organizations alike, they serve different objectives in private and government settings. Whereas the private firm seeks metrics that make it more profitable, organizations like the DoD and the Navy seek metrics that will help them achieve non-monetary goals, notably providing for national defense.

While the supra-ordinate objective of a private firm may differ from that of the Navy, new product development is critical to the success of both entities. Just as the private firm relies on its product development processes to make profitable products, the Navy relies on its procurement process to create and acquire the systems it needs to provide for national defense. Like any other organizational endeavor, the success of the product development process itself depends on the selection and exploitation of the right metrics. Whereas a firm's corporate

9

survival may depend on this, in the Navy's case, human life itself may hang in the balance. Therefore, if a private firm must "measure and control the product development process to ensure that the end product is exactly what the customer wants" (Majumder 2000), it is even more essential that the Navy measure and control its acquisition process to ensure that the end product is exactly what the war-fighter needs and what the Navy can afford to operate and support.

The Center for Innovation in Product Development at MIT has produced a continuing stream of research into metrics and incentives for, until recently, commercial product development. The thrust of this thesis research is to apply some of these latest product development ideas to Navy acquisitions to ensure that the Navy develops and procures the most cost-effective systems possible.

1-2   *Thesis Overview*

Chapter 2 describes the context of this research, the need to reduce government oversight in the acquisition process and the need to control total life cycle costs from the very conception of new systems.

Chapter 3 outlines the theory behind the Metrics Thermostat and the equations used to assign incentive weights to metrics. Chapter 3 also discusses the suitability of the theory to defense acquisitions.

Chapter 4 describes the process of data collection in Navy acquisitions and support organizations as well as the metrics chosen for this research.

Chapter 5 presents the statistical methodology used in this research and the results of the statistical analysis.

Chapter 6 summarizes the results of this research and suggests directions for further research.

This research comes at a period of transition in the defense procurement world; an on-going period of change precipitated by the end of the Cold War. While no one would lament the Cold War's passing, it left the DoD with a two-fold obstacle to maintaining readiness. First, in the ensuing years, the US government greatly reduced defense spending in pursuit of the much anticipated "peace dividend." As of 1996, defense spending had declined by 40 percent from the mid 1980's and weapons procurement had declined by 70 percent (Perry 1996). To compound the problem of maintaining readiness with reduced funding, the DoD's procurement process was not suited to the new lean environment. One 1992 study "calculated that the management and control costs associated with the DoD acquisition process were about 40% of the DoD acquisition budget, as compared to 5% to 15% for commercial firms" (Perry 1994). Thus, the Cold War's aftermath left the DoD with less funding to procure new technologies and a high-overhead, intensely bureaucratic method of doing so.

Recognizing the DoD's predicament, then Secretary of Defense William Perry issued a mandate for change in February of 1994. While there have been many initiatives over the years to make the acquisition process more cost effective, the "Perry mandate appears to have started the most recent efforts to reduce DoD costs" (R-TOC 2000). The mandate identified many problems with the defense acquisition system. Several of these problems shared a common source: the "complex web of laws, regulations, and policies" governing the defense acquisition system (Perry 1994). Over the years, the DoD had adopted detailed military specifications and oversight policies and required them of its suppliers and contractors. These detailed specifications and oversight policies were intended to ensure quality, but resulted in an "excessively high cost of doing business . . . due to telling contractors how to do the job as opposed to providing performance specs" (DSMC 1997). In addition to incurring high overhead costs, the detailed specifications and oversight policies also increased acquisition cycle times and in many cases, DoD systems were (and sometime still are) technologically obsolescent by the time they were (are) fielded (Perry 1994). As a partial remedy to these problems, the mandate called for the following changes (in addition to others not mentioned here):

- Move from rigid rules to guiding principles.

- Get bureaucracy out of the way.

- Foster competition, commercial practices, and excellence of vendor performance (increase reliance on the commercial marketplace).

Thus, in recent years, a large portion of the effort to reform the acquisition system has focused on relaxing military specifications and oversight. The new acquisition philosophy seeks not to dictate to suppliers and contractors exactly how to make new systems (i.e. military specs), but rather to establish performance specifications, allowing maximal leeway for achieving them.

While the DoD has been streamlining the procurement process and giving more freedom to contractors and suppliers, cuts in defense spending have made a priority of controlling the total ownership cost (TOC), or total life cycle cost, of new systems as they are being procured. In addition to calling for a streamlining of the procurement process, the Perry mandate required that the DoD "Adopt business processes characteristic of world-class customers and suppliers" (Perry 1994). According to the Reduction of TOC (R-TOC) Working Group at the Institute for Defense Analysis (IDA), "This point is not simply a re-statement that the DoD must procure items less expensively. Rather, the point is a call for DoD to mimic businesses that are driven by the 'bottom-line' metric. That metric ties the quality of the equipment to the *total cost of ownership of the system*" (R-TOC 2000, emphasis added). The IDA suggests that prior to the mid 1990's, "The major thrust of [efforts to improve defense acquisitions] was in the area of reducing acquisition costs." Since unit costs were much easier to track than operating and support costs (O&S costs), "intense focus remained on acquisition costs, and attempts to control life cycle costs were minimal" (R-TOC 2000).

Ironically, however, the acquisition cost of a system is quite literally "the tip of the ice berg" in terms of TOC (NPS 1999).

Figure 2-1 – Acquisition Cost vs. O&S Cost (NPS 1999)
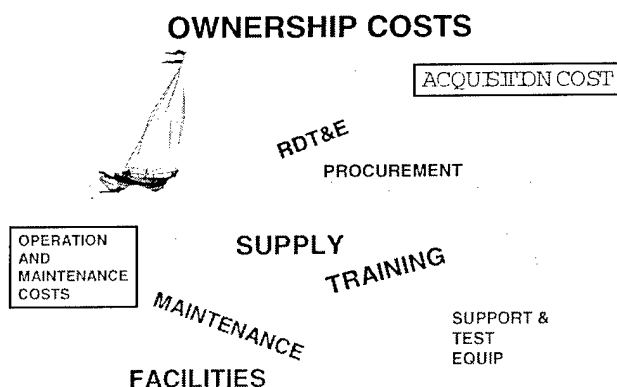
As a percentage of total life cycle cost, the acquisition cost may vary from system to system. However, O&S costs almost always determine the lion's share of TOC, in some cases up to 75% (Blanchard 1998).

Figure 2-2 –Life Cycle Costs by Category and Proportion (OSD-CAIG 1992)

LIFE CYCLE COST

OPERATING AND SUPPORT COST

DISPOSAL COST

INVESTMENT COST

RESEARCH AND DEVELOPMENT COST

CONCEPT EXPLORATION/ DEFINITION PHASE

DEMONSTRATION/ VALIDATION PHASE

ENGINEERING/ MANUFACTURING DEVELOPMENT PHASE

PRODUCTION AND DEPLOYMENT PHASE

OPERATIONS AND SUPPORT PHASE

DISPOSAL PHASE
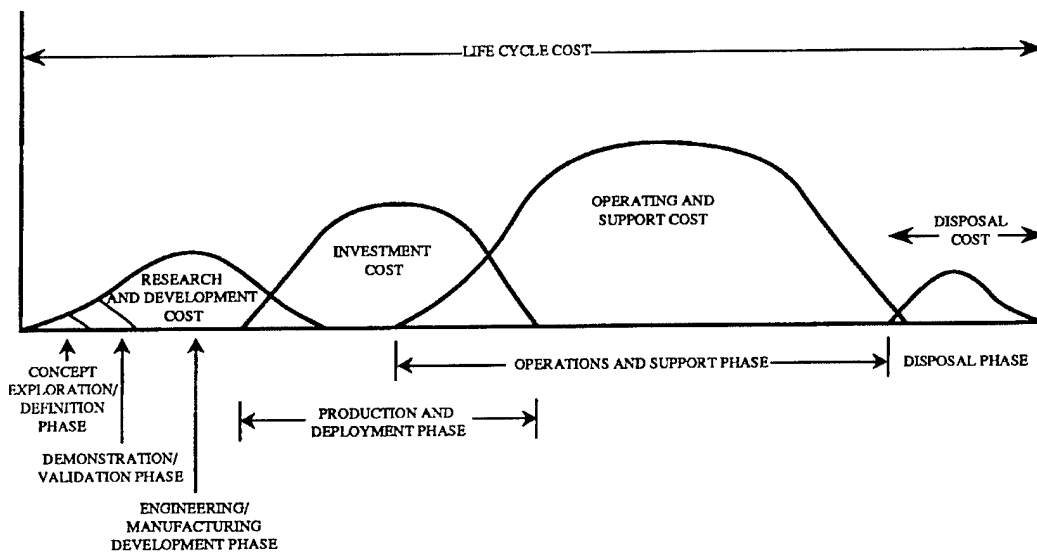
There are examples for which the figures are even more skewed. Thus far, O&S costs have accounted for 78 and 84 percent, respectively, of the life cycle costs of the F-16 and M-2 Bradley Fighting Vehicle (OSD-CAIG 1992).

Thus, O&S costs typically account for the largest portion of TOC. Moreover, "one often finds that a significant portion of this cost stems from the consequences of decisions made during the early phases of advance planning and conceptual design" (Blanchard 1998). Therefore, the opportunity to reduce O&S costs, and therefore, TOC, is greatest in the early stages of design and development (Blanchard 1998).

One of the most important initiatives addressing the DoD's need to "reduce life-cycle costs early in the acquisition process" is the policy known as Cost As an Independent Variable (CAIV) (Kaminski 1995). The policy was first proposed in a 1995 memorandum by then Undersecretary of Defense for Acquisition and Technology Dr. Paul Kaminski (R-TOC 2000). According to the Defense Systems Management College (DSMC), "CAIV is a new DoD strategy that makes total life-cycle cost, as projected within the new acquisition environment, a key driver of system requirements, performance characteristics, and schedules" (DSMC 1997). At the heart of this new strategy are "performing timely cost-performance trades" and "aggressively managing programs to meet those objectives, thus making [total life cycle] cost a major driver" (Kaminski 1995). This represents a marked departure from the old procurement system in which "requirements, performance, and sometimes schedule [drove] costs" (DSMC 1997). Thus, the new acquisition strategy emphasizes making the right life cycle cost - performance tradeoffs in order to achieve the best possible performance within budgetary constraints.

In order to implement this new acquisition strategy, Under Secretary Kaminski's CAIV working group called for new incentives for achieving cost objectives. The working group stated that, "Current practices frequently provide little or no industry incentive to reduce long-term costs to the government" (Kaminski 1995, Attachment 2). Furthermore, the CAIV working group stated that

> We need credible models to track projected unit production costs and O&S costs through development and into production . . . *Since O&S costs are not easily measurable in the early stages of the acquisition process, incentives to reduce O&S costs may require a (validated) model that relates specific design parameters* [i.e. metrics] *to measurable and predictable O&S costs* (emphasis added).

This last statement by the CAIV working group captures precisely what this research purposes to accomplish. The DoD seeks to manage and control the procurement process in such a way as to balance the TOC of new systems with effectiveness from their very conception. At the same time, there has been great effort to reduce the amount of bureaucracy and oversight in the acquisition process. The times call for an approach that will melt the iceberg in Figure 2.1 from the bottom up by reducing O&S costs while simultaneously avoiding the overhead and delays of micro-management. It is the premise of this research that the Navy can reduce the TOC of new systems by tying contract incentives to metrics that are valid predictors of O&S costs. At the same time, exploiting these metrics may help the Navy further reduce the amount of overhead in its acquisition process.

Chapter 3: Theory: The Metrics Thermostat

3-1 Overview

The previous chapter described the emergence of two major currents in contemporary defense procurement thought:

- To the maximum extent possible, the government must avoid telling contractors exactly how to build systems. Rather, the government should provide performance requirements and leave the details of achieving them to the contractor.

- The defense acquisition system must proactively manage and control the TOC of new systems, early in their development.

This research intends to demonstrate that a methodology developed at MIT by Professor John Hauser can help the US Navy achieve these two somewhat conflicting goals. Though Prof. Hauser initially developed this theory, called the Metrics Thermostat, with a commercial product development (PD) context in mind, Keith Russell has recently applied it with great success to a US Air Force maintenance organization (Russell 2000). For applications of the Metrics Thermostat in two large commercial firms, see LaFountain 1999 and Majumder 2000.

Section 3-2 gives a brief description of the mathematics and the steps for implementing the Metrics Thermostat, as derived for commercial PD. This description is somewhat condensed, as the emphasis of this thesis is primarily on the theory's application to Navy acquisitions. For a more detailed derivation of the equations used in the Metrics Thermostat, see John Hauser's paper, "Metric Thermostat."

Section 3-3 discusses the suitability of the theory to Navy acquisitions and the necessary adaptations in applying it to defense procurement.

3-2-1 The Context: Metrics and Agency Theory in Commercial PD

The Metrics Thermostat was conceived in the wake of the recent proliferation of information and information technology. This increasing abundance of information and information technology has changed the way companies develop new products. According to Maurice Holmes, former Chief Engineer at Xerox Corporation,

> This new product development vision is . . . about people working in a completely
> new way in an environment where traditional barriers to remote communication
> and collaboration are essentially eliminated. Its is about a major cultural reversal
> away from the era of exclusion, control, and co-location that product development
> managers worked so hard to build over the last 30 years (Keynote Address,
> PDMA 1999 International Conference).

With information technology breaking down the "barriers to remote communication and collaboration, there is a cultural shift to less centralized control" in commercial PD (Hauser 2000). This is the context for which the Metrics Thermostat was intended; a context in which "dispersed, self-directed, more autonomous teams are coordinated through common goals. We call those goals, metrics" (Hauser 2000).

More specifically, the Metrics Thermostat envisions a firm in which there are top level managers and subordinate product development teams that create the firm's new products. Top-level managers seek to maximize the profit from new products while the employees on the PD teams choose their actions so as to maximize their own best interests, as opposed to that of the firm (i.e. profitability). The Metrics Thermostat enables top-level management to align its PD teams' best interests with the firm's best interest, profit. Managers attempt to do so by providing their PD employees with the right incentives and rewards. If management chooses the incentives and rewards properly, then these incentives and rewards will lead the employees to choose actions that will increase the profitability of new products. This underlying concept is commonly referred to as agency theory, in which a principal (in this case management) contracts with an agent (the product development teams) to perform some task(s). Both parties act in their own best interests and the contract must be structured so that the incentives and rewards to the agent align his/her best interests with that of the principal (management).

As mentioned in the introductory chapter, metrics inevitably become the basis for incentives and rewards, whether explicitly or implicitly. Metrics don't just measure, they become incentives in their own right, for "what gets measured, gets done," as the age old adage implies. Choosing the right metrics, therefore, is critical to success. Consider, for example, the following examples. At Xerox, Chief Engineer Maurice Holmes successfully implemented a plan to reduce time-to-market (TTM) by a factor of 2.5 (Hauser 2000). On the other hand, a poor choice of metric(s) can lead to disastrous results. Gibbons (1997) cites several examples:

> At the H.J. Heinz Company, division managers received bonuses only if earnings
> increased from the previous year. The managers delivered consistent earnings

growth by manipulating the timing of shipments to customers and by prepaying for services not yet received, both at some cost to the firm (Post and Goodpaster, 1981). At Bausch & Lomb, the hurdle for bonuses was higher, often entailing double-digit earnings growth. Again, managers set their targets in ways that were not obviously in the best long-run interests of the firm (e.g., over a half-million pairs of "sold" sunglasses were discovered in a warehouse in Hong Kong; Maremount 1995).

Firms must not only choose good metrics, they must also determine a relative emphasis to place on them. For example, reduced TTM and greater customer satisfaction both (generally) make a product more profitable and firms typically measure both. However, greater effort to increase customer satisfaction will usually increase TTM. Likewise, effort to decrease TTM may require investing less time and effort in customer satisfaction. Companies must, therefore, determine how to prioritize among sometimes competing metrics in order to maximize profit.

Determining the relative emphasis, or weight, to assign PD metrics in order to maximize profit is precisely the objective of the Metrics Thermostat. The Metrics Thermostat assesses the value of a metric relative to other metrics. The following passage summarizes the core concept of how the Metrics Thermostat determines a metric's value to the firm:

> A metric is often defined as something that can be precisely measured, but this definition may mislead modern organizations into misuse of their metric systems. A precisely measured metric may be precisely wrong where a harder to measure metric may be vaguely right. Perhaps management wants to know how productive their sales force is. They may be precisely able to measure the number of telephone calls sales people make each day (a precise but less accurate measure of productivity). Alternately, they may choose to conduct a survey of telephone customer satisfaction (a less precise but, perhaps, more representative metric for worker productivity). So, the value of a metric can be determined by two characteristics: its measurement precision and [the closeness of] its association to its target concept (Russell 2000).

The Metrics Thermostat assumes that the firm has already chosen its metrics and then provides a basis by which the firm prioritizes, or weights, them and thereby gives incentives to its employees.

3-2-2 Theory Formulation

The following section is a condensed explanation of the theory and implementation of the Metrics Thermostat. The explanation is based on John Hauser's paper "Metrics Thermostat." For a more rigorous treatment, refer to "Metrics Thermostat" (Hauser 2000).

3-2-2-1 Notation and Assumptions

The reader may find the notation and variables in the following paragraphs somewhat difficult to track. However, the gist is rather simple. The PD team chooses to perform a large number (K) of individual actions in the PD process. Rather than dictating these individual actions, management measures a smaller number (n) of metrics and rewards the PD team (monetarily and/or non-monetarily) to the extent that the actions taken by the PD team increase (or decrease or maintain) the levels of these metrics. Thus, management determines and measures strategic metrics it believes to correlate with profit, while the PD team chooses the individual actions to take in the design process. These actions, in turn, determine the profitability of new products. Additionally, the PD team's actions determine the level of effort exerted towards each metric, and therefore, the rewards to the PD team. The key is in determining how much reward to give for each metric in order to maximize the net increase in profit.

- For the sake of simplicity, we assume for the moment, that the firm has only one PD team. We will relax this assumption later, since this is not the case in most firms.

- We denote the firm's profit with the Greek letter π.

- Profit (π) is a function of the actions the PD team takes. There are a myriad of individual tasks the PD team must perform in the design process. Examples include using a house of quality to increase customer satisfaction or applying some platform reuse methods to reduce TTM (Hauser 2000). These individual actions we will denote as a_k for k = 1 to K.

  o There is an inherent cost to these actions, $c(a_1, a_2, \ldots a_K)$. The PD team incurs this cost and the details of these costs are not visible to management. Cost is defined in the most general sense. It may be monetary or non-monetary. It is the combined direct cost the PD team incurs by its actions as well as the opportunity costs of these actions. To reiterate, the PD team incurs this cost, not the firm.

- There is a status quo, or current operating point for the firm on the profit curve. The current operating point is defined by all the actions that the PD team has taken in the development of recent products.

- Rather than dictating the teams' individual actions, management chooses metrics, m_i for i = 1 to n, where n is typically much smaller than K.

  o We associate with each metric, m_i an unobservable level of effort, e_i^a that the PD team exerts towards that metric. The actions that the PD team chooses, $\{a_1, a_2, \ldots a_K\}$ determine the levels of the efforts, e_i^a that the PD team exerts.

In the context of the previous example, the team might choose to make a house of quality (an action), thereby increasing the level of effort it exerts toward some measure (metric) of customer satisfaction. Therefore, each metric, m_i is a function of the efforts, e_i^a that the PD team exerts, (i.e. $m_i = m_i (e_i^a)$). Profit, in turn, can be expressed as a function of the metrics, (i.e. $\pi = \pi (m_1, m_2, \ldots, m_n)$).

o The status quo for the firm is determined collectively by the current levels of effort e_i^0. Together, these efforts, $\{e_1^0, e_2^0, \ldots e_n^0\}$ represent the firm's current operating point on the profit curve, π. They represent the level of effort that the PD team has been exerting toward each metric.

- Alternately, since profit can be expressed as a function of the metrics, the firm's current operating point can also be represented by the current level of each metric, m_i^0. This current operating point corresponds to the current level of profit, π^0.

o We denote any incremental efforts the team takes to change the initial operating point as e_i. We can then rewrite the PD team's cost function: $c(a_1, a_2, \ldots a_K) \rightarrow c(e_1^0+e_1, e_2^0+e_2, \ldots, e_n^0+e_n)$, or more simply, $c^0(e_1, e_2, \ldots, e_n)$.

- Management's own measures of the firm's PD metrics contain some error, or noise. Thus, the measure of each metric, $\tilde{m}_i$ can be written as the actual value of the metric m_i plus some "white noise:" $\tilde{m}_i = m_i(e_i^a)+error_i$ (where the error associated with the measurement of each metric has zero mean and is normally distributed with variance σ_i^2).

- The firm provides its PD team with rewards based on the observed incremental changes in the metrics about the current operating point.

o The Metrics Thermostat assumes that the firm provides rewards to the PD team as a weighted linear sum of these observed incremental changes. Thus, the rewards it provides to the PD team can be written: $rewards=w_0+w_1 \tilde{m}_1+w_2 \tilde{m}_2+\ldots+ w_n \tilde{m}_n$.

- The term w_0, represents a base salary or other benefit given to the team.

- The weights, w_i, are the relative emphasis, or reward placed on each metric m_i. As Russell notes, "most organizations do not pay employees based on a set of metrics (although many sales forces pay on commission). Instead, management signals employees what the organization believes is important by establishing pay raises, providing bonuses, and giving other incentives based on the team's ability to [increase, decrease, or maintain the level of] these metrics (Russell, 2000).

- Recall that the $\tilde{m}_i$ are random variables, each with variance σ_i^2. Essentially, this means that management's perception, or measurement of each metric is not exact. An employee may exert more effort towards a metric than management perceives. Alternately, an employee may exert

19

less effort towards a metric than management perceives. The Metrics Thermostat assumes that the PD team is risk averse, preferring less risky metrics (i.e. metrics with smaller variances) given the same incentive or weight. In other words, given two metrics with the same reward, the PD team will work harder on the one that is less risky (i.e. more precisely measured by management). By the same principle, a risk-averse investor will prefer the stock that is less risky, given two stocks that yield the same expected return.

3-2-2-2 Solving for the Optimal Weights

The PD team chooses its actions (or efforts, $\{e_1^0, e_2^0, \ldots e_n^0\}$) in such a way as to maximize what is referred to as the team's certainty equivalent. The team's certainty equivalent (CE) is the rewards to the team minus the cost the team incurs for its efforts. Since the team is risk averse, it discounts the reward for each metric in proportion to its risk, or, variance. Therefore, the PD team maximizes its CE:

(3.1) $\quad CE \approx w_0 + w_1 m_1 + w_2 m_2 + w_n m_n - c^0(e_1, e_2, \ldots, e_n) - \frac{1}{2}r\, w_1^2\, \sigma_1^2 - \frac{1}{2}r\, w_2^2\, \sigma_2^2 - \ldots - \frac{1}{2}r\, w_n^2\, \sigma_n^2$

Note that the term r denotes the degree to which the PD team is risk averse.

After solving the PD team's optimization problem, the Metrics Thermostat chooses a weight (or relative importance) w_i for each metric m_i in such a way as to maximize the firm's net increase in profit (the incremental gains in profit from the efforts the team exerts toward each metric minus the rewards paid to the team based on measurements of the metrics):

(3.2) $\quad$ net profit $\approx \pi^0 + \sum_{i=1}^{n} \frac{\partial \pi}{\partial e_i^0} e_i - (w_0 + w_1\, \tilde{m}_1 + w_2\, \tilde{m}_2 + \ldots + w_n\, \tilde{m}_n)$

After solving the PD team's optimization problem, the results are used to maximize the firm's net increase in profit. The solution to this problem yields the weights that the firm places on each metric. The optimal weight for each metric is given by the equation:

(3.3) $\quad w_i^* = \dfrac{\left(\dfrac{\partial \pi}{\partial e_i^0} \Big/ \dfrac{\partial m_i}{\partial e_i^0} \right)}{\left[1 + \left(r \dfrac{\partial^2 c^0}{\partial e_0^{02}} \right)\left\{ \sigma_i \Big/ \left(\dfrac{\partial m_i}{\partial e_i^0} \right) \right\}^2 \right]}$

3-2-2-3 An Intuitive Explanation of Terms

By decomposing Equation 3.3 into its three main terms, one can obtain a more intuitive understanding of the Metrics Thermostat's weighting scheme.

The numerator of Equation 3.3 is referred to as the metric's leverage.

20

Term 1: Leverage: $\left(\dfrac{\partial \pi}{\partial e_i^{\,o}} \middle/ \dfrac{\partial m_i}{\partial e_i^{\,o}}\right)$

The leverage term captures the metric's marginal effect on profit. Leverage represents the incremental change in profit per unit increase in the level of the metric. All other things being equal, the firm should weight metrics in direct proportion to their effect on profit.

Term 2: The Noise-to-Signal Ratio: $\left\{\sigma_i \middle/ \left(\dfrac{\partial m_i}{\partial e_i^{\,o}}\right)\right\}^2$

As its name implies, the Noise-to-Signal Ratio (NSR) represents the "noisiness" of the metric. The numerator of this term is the metric's standard deviation, which represents the magnitude of the metric's error normalized by the scale of the metric. The SNR appears in the denominator of Equation 3.3 since the more noisy the metric (i.e. the more prone to error the metric is), the less the firm should weight it.

Term 3: The Risk-Effort Aversion Term: $\left(r\dfrac{\partial^2 c^o}{\partial e_i^{\,o2}}\right)$

The other term in the denominator represents the product of the PD team's aversion to risk (r) and its aversion to effort ($\partial^2 c^o / \partial e_i^{\,o2}$). All other things being equal, the more risk and the more effort to increase a metric costs the PD team, the less emphasis it will receive.

Terms 1 and 2: The Tradeoff of Accuracy and Precision: $\dfrac{\left(\dfrac{\partial \pi}{\partial e_i^{\,0}} \middle/ \dfrac{\partial m_i}{\partial e_i^{\,0}}\right)}{\left\{\sigma_i \middle/ \left(\dfrac{\partial m_i}{\partial e_i^{\,o}}\right)\right\}^2}$

The ratio of terms 1 and 2 captures the tradeoff between using precise measurements and using accurate measurements. "'Soft' metrics, such as customer satisfaction [as measured by a survey], might have higher leverage than 'hard' metrics, such as the number of defects reported. The tradeoff is that the soft metrics will have [higher NSR's]" (Hauser 2000). Thus, one assesses a metric's value according to how closely it corresponds to the construct that it is attempting to measure and also the error of the metric's measurement.

3-2-2-4 The Empirical Form of Equation 3.3

Equation 3.3 gives the weight for each metric, however, empirical use of the formula requires two major adaptations.

First, we have assumed thus far that the firm has only one PD team. The weights for each metric (w^*_i) are meant to quantify the implicit emphasis a firm places on a metric by the culture it sets and the leadership style it adopts (in addition to direct monetary incentives). As a practical matter, management cannot set a different corporate culture or adopt a different leadership style for individual PD teams. At the same time, some of the terms in Equation 3.3 may not be homogeneous throughout the firm, especially if the firm has divisions that make very different products and/or have very different cultures. If, however, there is sufficient homogeneity within these divisions, then Equation 3.3 may be tailored to each of the firm's major divisions (Hauser 2000).

Second, implementing Equation 3.3 directly is not practical since management does not know the exact values of the Leverage, NSR, and Risk-Effort Aversion terms. Rather, management must estimate these quantities statistically.

Using a tangent hyperplane approximation to the profit curve, Hauser shows that the numerator of Equation 3.3 (i.e. Leverage) can be estimated with a multiple (linear) regression coefficient, $\hat{\lambda}_i$. This term, $\hat{\lambda}_i$ is the regression coefficient of the metric, m_i when profit is regressed on all the metrics. The denominator of Equation 3.3 may be estimated by a quantity referred to as the risk discount factor (RDF). The RDF is the "amount by which a team will discount the real, risky rewards [of a metric] relative to a situation where the rewards can be guaranteed" (Hauser 2000). RDF measures the "net effect of risk aversion, effort aversion, and the [NSR] of a metric" (Hauser 2000). For a detailed explanation of RDF, refer to Hauser 2000.

Equation 3.4 gives the empirically measurable form of Equation 3.3:

$$3.4 \qquad \hat{w}^d_i = \frac{\hat{\lambda}_i}{1 + 2RDF_i}$$

Note that the superscript d denotes that the firm estimates these quantities within each sufficiently homogeneous division of the firm. The firm must exercise its judgment as to what constitutes a sufficiently homogeneous division. As a guiding principle, management may wish to consider the extent to which the three terms explained in Section 3-2-2-3 differ within the major units of the organization.

3-2-3 Implementing the Metrics Thermostat Process

Thus far, we have concentrated on the Metrics Thermostat's method for prioritizing metrics. The Metrics Thermostat is an iterative process, not just a "one-time" method for evaluating the relative emphasis to accord a set of metrics. Equation 3.3 does not maximize profit, rather it specifies a weighting system for metrics that will induce PD teams to make incremental efforts in

the direction of steepest ascent up the profit curve. Once the firm sets the weights associated with each metric, the culture of the firm will change and employees will adjust their actions in accordance with the new reward system. This, in turn will cause the company to move "up" the profit curve. If the company takes no further action to measure the impact of these changes, then it may "overshoot" on some metrics. The reader may wish to consider an example in which a person wishes to walk up a hill. If a person sets out in the direction of steepest ascent and walks too far in that direction, then he/she will eventually pass the top of the hill and begin descending down the backside of the hill. Similarly, if management does not periodically reassess its current operating point and the direction in which the firm's PD is heading, then it risks overshooting the top of the profit curve. Just as a thermostat constantly measures room temperature and tells an HVAC system to heat, cool, or stand by, the MT repeatedly measures the firm's current operating point against a set of metrics and profit to determine the direction in which the firm should head to maximize the incremental change in profit.

Hauser (2000) outlines the following seven-step process for practical implementation of the Metrics Thermostat:

1. Identify a set of PD projects that follow approximately the same culture.

2. Identify the metrics by which the firm is managed.

3. Use the firm's documentation to obtain measures of the metrics, and profit in the last Y years (typically Y=5).

4. Use multiple regression to obtain estimates of leverage ($\hat{\lambda}_i$) for each metric.

5. Use survey measures to obtain the Risk Discount Factor (RDF$_i$) for each metric.

6. Use Equation 3.4 to calculate ($\hat{w}^d_i$) for each metric. Increase or decrease the emphasis on each metric as indicated.

7. Return to step 3 periodically to update ($\hat{w}^d_i$) Optimality is reached when ($\hat{w}^d_i$) = 0, but periodic monitoring enables the system to adjust to environmental changes.

23

Thus far in this chapter, we have outlined the foundational theory of the Metrics Thermostat and its implementation in commercial (for-profit) product development. We now turn our attention to the suitability of the Metrics Thermostat to naval acquisitions.

It is appropriate at this point to consider some of the differences between defense and commercial PD and the resultant modifications necessary for applying the Metrics Thermostat to naval acquisitions. The different objectives of naval acquisitions and commercial PD constitute the first, and most obvious difference between them. In commercial PD, profit is the ultimate goal of the organization, whereas other goals such as readiness, effectiveness, and reduced TOC replace profit in the Navy context. This poses no problem as long as the supra-ordinate goal replacing profit is quantifiable and measurable. For example, in applying the Metrics Thermostat to an Air Force maintenance organization, Keith Russell used F-16 Mission Capable rates as one of the organization's supra-ordinate goals. This research focuses on reducing the O&S cost of new systems as the supra-ordinate goal (replacing profit).

A second difference, following almost as a corollary to the first, is that defense and commercial PD will use somewhat (but not entirely) different subordinate PD metrics. Operational Availability (Ao), for example, is an important metric to the Navy, but it is not typically used in commercial PD. Some metrics, however, are strikingly similar to those used in commercial PD. For example, the Navy may measure mean-time-between-failure (MTBF), which is analogous to the defect rate of a commercial product. Again, differing subordinate metrics pose no problem to implementing the Metrics Thermostat as long as they are quantifiable and measurable. Chapter 4 gives a lengthy description of the particular metrics included in this research.

The last important difference between Navy and commercial PD considered here is that the players involved are somewhat different. In this research, the US Navy (NAVSEA, or perhaps the DoD, or even Congress for large acquisitions) becomes the principal and the defense contractors who design and build Navy systems become the agents working for the principal. Defense contractors are perhaps more autonomous of the Navy than the PD divisions within a firm, so the relationship of NAVSEA and defense contractors may be somewhat different than that between management and PD teams within the same firm. However, the underlying principle of a principal contracting with an agent, each maximizing their own best interests, accurately describes the relationship of the Navy and defense contractors. The exact nature and type of the rewards to the agent (defense contractors) may change somewhat, but rewards and incentives remain integral to the relationship between the Navy and defense contractors.

Though the Metrics Thermostat was developed for commercial PD, it is the argument of this research that is may prove very useful in the naval acquisitions process. The beginning of this chapter recapitulated two important issues in defense procurement, repeated here:

- To the maximum extent possible, the government must avoid telling contractors exactly how to build systems. Rather, the government should provide performance requirements and leave the details of achieving them to the contractor.

- The defense acquisition system must proactively manage and control the TOC of new systems, early in their development.

The Metrics Thermostat provides for both goals. Rather than dictating the contractors' individual actions, the Metrics Thermostat allows the Navy to manage with metrics. The agency theory aspect of the Metrics Thermostat ties the contractors' rewards to achieving the second goal of reducing TOC of new systems. The Metrics Thermostat is precisely what the CAIV working group prescribed when it called for "a (validated) model that relates specific design parameters [i.e. metrics] to measurable and predictable O&S costs."

- This page intentionally left blank -

4-1   *Overview*

The first three steps of the Metrics Thermostat (repeated below) tailor the general theory of Chapter 3 to an organization's specific goals, metrics, and available data:

1. Identify a set of PD projects that follow approximately the same culture.

2. Identify the metrics by which the firm is managed.

3. Use the firm's documentation to obtain measures of the metrics, and profit in the last Y years (typically Y=5).

In this chapter, I describe how I applied these steps to the organizations responsible for procuring and maintaining Navy systems. Section 4-2 provides a narrative describing the "data and people trail" I followed in accomplishing steps 1 through 3. I describe the major steps along this path and the major influences and insights that came along the way. I describe important observations and deductions drawn from the data-people trail I followed in the Navy acquisitions and technical support community.

In Section 4-3, I focus on step 3 of the Metrics Thermostat and describe the major sources of data available for this research. The observations and deductions of Section 4-2, in conjunction with the available documentation described in 4-3, provided the basis for the metrics selected for this research.

The individual metrics are listed in detail in Section 4-4.

As listed in Section 4-1, the first three steps of the Metrics Thermostat suggest a simple, sequential process for tailoring the Metrics Thermostat to a specific organization. In this research, however, steps 1, 2, and 3 were not performed sequentially, but rather, simultaneously and iteratively. At this point, I rephrase these steps as questions that lead me down a data and people trail in the Navy.

1. What constitutes a data point (i.e. PD project) for the Navy?

2. What metrics are appropriate for Navy PD?

   a. What metric (or metrics) replaces profit in the Navy PD context as the over-all goal of naval acquisitions?

   b. What are the lower-level metrics that drive the over-all, supra-ordinate goal(s) of naval acquisitions?

3. What data and documentation does the Navy keep regarding these data points and metrics?

(I must also add that I did not embark on the data trail alone, but with Lieutenant Commander (LCDR) Carl Frank, (then) a student at the Sloan School of Management whose thesis work (Frank 2000) also addresses implementing the Metrics Thermostat to Navy acquisitions, and also with the help of Mr. Thomas Kowalczyk of the Office of Naval Research (ONR), who introduced us to several contacts within the Navy.)

Previous applications of the Metrics Thermostat in the commercial sector had used PD projects such as copy machines and automobiles as data points. At the outset of this research, it was not clear what the corresponding data point should be for applying the Metrics Thermostat to Navy PD. We needed to choose data points that were relatively similar (i.e. follow a similar culture), yet sufficient in number, as previous applications had suffered from a paucity of data. In our earliest meetings with Navy representatives, we identified two possible candidates, ships and shipboard systems. Subsequent meetings at NAVSEA revealed that the problem would be much more tractable if we used shipboard systems as data points. If we had chosen ships as our data points, the need for homogeneity may have limited us to ships within a particular ship class. However, even within a ship class, individual ships are somewhat unique due to the differing manners in which their crews operate and maintain them and the differing configurations of the equipment on board (depending on when and where the ship was made). By choosing shipboard systems, we did not completely eliminate these problems since different crews operate and maintain each system and many systems exist in different configurations throughout the fleet. Nonetheless, our earliest discussions with Navy representatives suggested that shipboard systems were more suitable for the study than ships. Moreover, the number of shipboard systems in the fleet far exceeds the number of ships within a particular ship class. Therefore, choosing shipboard systems afforded us access to more data. Furthermore, we were encouraged to use shipboard systems as data points as we learned that the Navy maintained similar metrics for a

large number of systems, despite the fact that many of these systems perform very different functions. Thus, we chose shipboard systems over ships as our data points.

Once we had decided to focus on shipboard systems, we still needed to determine exactly which systems to use. This did not become clear until we had spent considerable time addressing the second and third questions mentioned above.

In our quest for the appropriate supra-ordinate goal(s) (replacing profit), subordinate metrics, and data sources, LCDR Frank and I met with representatives from numerous agencies within the DoD and Navy. The following are some of the organizations we consulted in the early stages of this research:

- The Office of Naval Research
- The Director of the Research and Development Group at NAVSEA
- The NAVSEA Command Metrics Coordinator
- The Defense Advanced Research Projects Agency
- The Head of the Systems Supportability Engineering Branch, Naval Undersea Warfare Center
- The Naval Post Graduate School, Department of Systems Management
- The Navy Manpower Analysis Center
- The Top Management Attention/Top Management Issues Program (Atlantic)
- COMNAVSURFLANT
- The Institute for Defense Analysis
- The Naval Surface Warfare Center, Port Hueneme Division
- The Open Systems Joint Task Force
- The Affordability Through Commonality Program

The need for affordability and reduced TOC of new systems emerged as a recurring theme in many of the discussions with representatives from these groups. It soon became apparent that some measure of TOC would become one of our supra-ordinate metrics, replacing profit in e Equation 3.3. (This only stands to reason, in light of the events described in Chapter 2.)

In addition to reduced TOC, other subordinate goals and desired outcomes emerged from a quasi-consensus among the groups consulted. These include:

- Reduced manning and training requirements
- System Reliability, Maintainability, and Availability (R, M,& A)
- System Supportability
- Ease of technology insertion/refreshment, or "upgradability"
- Openness of System Architecture
- Inclusion of COTS/NDI technology
- Software Support and Maintenance Costs

It should be noted that not every group agreed on the relative importance of the above goals and a few even suggested that some of the goals were actually getting too much emphasis.

Moreover, in some cases, the groups differed as to the definitions and meanings of some of these terms. Nevertheless, the fact that these same concepts consistently surfaced in meetings with the various groups we visited revealed an implicit consensus on the importance of these concepts in Navy PD (though not always on their definitions and relative importance).

Visits with analysts from the Top Management Attention/Top Management Issues program (TMA/TMI) in Norfolk, VA proved especially helpful in identifying some of the specific shipboard systems available as data points. Moreover, our interactions with the TMA/TMI program exposed us to many of the metrics that the Navy tracks on its shipboard systems as well as potential sources for data. The TMA/TMI program uses a set of metrics "to identify systems/equipments as candidates for consolidated action to improve performance" (TMA/TMI website). Like the Metrics Thermostat, the TMA/TMI program uses a set of metrics to recommend incremental actions to improve the performance (and cost) of Navy systems. The TMA/TMI metrics are primarily related to how often a system fails, and how much time and money are required for corrective maintenance. These metrics were an excellent starting point for determining some of the specific metrics to use in applying the Metrics Thermostat to Navy PD.

In addition to TMA/TMI, the Port Hueneme Division of the Naval Surface Warfare Center (PHD NSWC) provided a very helpful resource, the "PHD NSWC Safety, Effectiveness, and Affordability Review Guide" (PHD NSWC 1999). The "PHD NSWC Safety, Effectiveness, and Affordability Review Guide" provides several R, M, & A metrics the Navy tracks as well as some important supportability and readiness metrics. In addition to providing some of the lower level, subordinate metrics important to the Navy, the SEA Review Guide also suggests a three dimensional measure of system effectiveness. According to the guide, a system's effectiveness is composed of:

- Capability of Performance (P_c): a measure of the system's inherent capability to perform the mission it was designed for as measured by the probability of the system achieving its mission assuming there is no equipment, computer program, or human error.

- Operational Availability (A_o): the "likelihood that, when required, a system is operating at a pre-defined level and for a sufficient duration of time to accomplish its mission."

- People Factor (P_p): "the probability of humans performing all of the necessary steps on time to properly set up and operate one or more systems and complete the mission."

Though TOC usually dominated our discussions of appropriate supra-ordinate metrics and goals for Navy acquisitions, we were also interested in some measure(s) of effectiveness for Navy systems. Those proposed by the SEA Review Guide were among the most objective and comprehensive we encountered. (Regrettably, the time frame of this research did not permit an application of the Metrics Thermostat to system effectiveness.)

With TOC figuring so prominently in our meetings throughout the Navy, it became necessary to find the best source of TOC data. Several people in our meetings referred us to the Naval Center for Cost Analysis, Visibility and Management of Operating and Support Costs (VAMOSC)

program. Navy VAMOSC maintains some of the most comprehensive cost data available for about 50 different Navy shipboard systems. With comprehensive cost data readily available, the systems for which VAMOSC kept data were the strongest candidates to become the data points for this research.

*Table 4-1 – Listing of VAMOSC Shipboard Systems**

Years	System
FY86-94	5"/54 CALIBER MK-42 GUN
FY86-99	5"/54 CALIBER MK-45 GUN
FY93-99	AN/BPS-15 SERIES(A-D) RADAR
FY97-99	AN/BPS-16 (V) RADAR
FY86-99	AN/BQQ-5 SONAR SYSTEM
FY97-99	AN/BQQ-6 SONAR
FY97-99	AN/BQS-15 SONAR DETECTING-RANGING SET
FY93-99	AN/BRD-7 AND 7A ELECTRONIC COUNTERMEASURE SET
FY93-99	AN/SLQ-25 AND 25A NIXIE TORPEDO COUNTERMEASURE SYSTEM
FY86-99	AN/SLQ-32 ELECTRONIC WARFARE SYSTEM
FY98-99	AN/SLQ-48(V) NEUTRALIZATION SYSTEM MINE
FY98-99	AN/SPS-40B RADAR
FY98-99	AN/SPS-40E RADAR
FY93-97	AN/SPS-40C/D/E
FY93-99	AN/SPS-48C RADAR
FY93-99	AN/SPS-48E RADAR
FY87-99	AN/SPS-49 RADAR
FY86-99	AN/SPS-55 RADAR
FY91-97	AN/SPS-64 (V) 3 AND 9 RADAR
FY98-99	AN/SPS-67 (V) 1 RADAR
FY98-99	AN/SPS-67 (V) 3 RADAR
FY91-97	AN/SPS-67 (V) 1 & 3 RADAR
FY99	AN/SQQ-32(V) 2 SONAR SET ADVANCED MINEHUNTING
FY99	AN/SQQ-32(V) 3 SONAR SET ADVANCED MINEHUNTING
FY98	AN/SQQ-32 SONAR SET ADVANCED MINEHUNTING
FY91-99	AN/SQQ-89 SURFACE ASW COMBAT SYSTEM
FY85-99	AN/SQS 53A SONAR
FY86-99	AN/SQS 56 SONAR
FY98-99	AN/SRS-1(V)SERIES SIGNAL DETECTION-DIRECTION FINDING SETS
FY93-99	AN/SYQ-20 ADVANCED COMBAT DIRECTION SYSTEM
FY93-99	AN/SYS-2 INTEGRATED AUTOMATIC DETECTION AND TRACKING SYSTEM
FY96-99	AN/URC-107 JOINT TACTICAL INFORMATION DISTRIBUTION SYSTEM
FY96-99	AN/USC-38 EHF SATCOM
FY97-99	AN/USG-1/2 COOPERATIVE ENGAGEMENT CAPABILITY (CEC)
FY96-99	AN/USQ-82(V) SHIP DATA MULTIPLEX SYSTEM
FY97-99	AN/WLQ-4 (V)/(V) 1 COUNTERMEASURE RECEIVING SET
FY97-99	AN/WLR-1H RADAR WARNING SYSTEM
F797-99	AN/WLR-8 (V) 2/ (V) 5 COUNTERMEASURE RECEIVING SET
FY85-95	ASROC LAUNCHING GROUP MK-16 AND FIRE CONTROL SYSTEMS
FY86-99	CLOSE-IN WEAPON SYSTEM MK-15

FY86-99	COMBAT CONTROL SYSTEM MK-1
FY97-99	COMBAT CONTROL SYSTEM MK-2
FY87-99	HARPOON WEAPON SYSTEM
FY86-99	LM 2500 GAS TURBINE ENGINE
FY93-99	MK 14 WEAPONS DIRECTION SYSTEM
FY97-99	MK 23 TARGET ACQUISITION SYSTEM (TAS)
FY87-99	MK 26 GUIDED MISSILE LAUNCHING SYSTEM
FY87-99	MK 41 VERTICAL LAUNCHING SYSTEM
FY97-99	MK 57 NATO SEA SPARROW SURFACE MISSILE SYSTEM
FY93-99	MK 74 MISSILE FIRE CONTROL SYSTEM
FY97-99	MK-75 76MM GUN OTO-MELARA
FY87-99	MK-86 GUN FIRE CONTROL SYSTEM
FY91-99	MK-92 FIRE CONTROL SYSTEM
FY96-99	MK-116 UNDER WATER FIRE CONTROL SYSTEM MOD 1/2 AND 4
FY86-99	MK-117 FIRE CONTROL SYSTEM
FY97-99	MK 118 UNDERWATER FIRE CONTROL SYSTEM (UFCS)
FY97-99	RAM MK 31 GUIDED MISSILE WEAPONS SYSTEM
FY96-99	TB-23 SUB TOWED ARRAY

Note, some systems were not included in the data set due to lack of available data, insufficient population size, or other considerations discussed in Chapter 5.

Two caveats merit discussion at this point. First, it should also be noted that some of the individuals consulted in the Navy, especially those in the engineering community, expressed reservations about the reliability of VAMOSC data. However, other forms of available cost data focus primarily on the annual maintenance costs of systems, which comprise only a fraction of TOC. The VAMOSC data is by far the most comprehensive, including much more than just maintenance costs (see Section 4-3 for a more detailed description of VAMOSC cost data). Even though the VAMOSC data may contain some reporting error (as all data inevitably does), its inclusiveness makes it more likely to provide a better estimate of a system's true TOC than other sources of cost data that only provide maintenance costs. Moreover, the other, less comprehensive sources of cost data are also likely to contain some reporting error, though perhaps less than the VAMOSC data if some of the Navy's engineers and program managers are correct.

The second issue to consider is that VAMOSC O&S cost data is not a 100% complete measure of TOC since it does not include all of the acquisition and disposal costs of a system. As illustrated in Chapter 2, however, these costs typically do not exceed 1/3 of a system's TOC, whereas O&S costs typically comprise the bulk of the TOC "iceberg." Finding the acquisition costs of the systems in Table 4.1 was attempted initially, but met with little success. Since many of these systems were purchased over the course of several years (in some case decades), determining the actual acquisition cost of a given system is quite complicated since the acquisition cost often varies over time. The Navy typically contracts to buy a certain quantity of a system over a specified period of time, but often buys more than one batch of the same system from different manufacturers at different times and at different prices. One might estimate a system's average acquisition cost, but in many cases, the documentation required to compute this has not survived the turnover in personnel and changes of location at the respective program offices. As a practical matter, the O&S cost data provided by Navy VAMOSC offers, in my

judgment, the best available measure of TOC. For this reason, the systems in Table 4.1 were chosen as the initial data points for this research. (Some of these systems were later excluded for reasons described in Chapter 5.)

With the first step of the Metrics Thermostat (and most of the second step) accomplished, we focused our efforts on identifying specific metrics to measure the important concepts and goals that came out of our meetings in the Navy acquisitions and support organizations, as well as sources for these metrics.

Our journey down the data and people trail identified at the most abstract level, the supra-ordinate goals and metrics for implementing the Metrics Thermostat to Navy acquisitions as well as the general subordinate goals and concepts that drive them.

- Supra-Ordinate Goals (Outcomes):

  o Operating and Support Cost

  o System Effectiveness*

- Subordinate Driving Concepts (Outcome Drivers):

  o Manpower and Training Requirements

  o System Reliability

  o System Maintainability

  o System Availability

  o System Supportability*

  o Capability of Performance*

  o The "People Factor"

  o Ease of Technology Insertion/Refreshment, or "Upgradability"

  o Openness of System Architecture

  o Inclusion of COTS/NDI Technology

  o Software Support and Maintenance Requirements

The next step was to determine the specific measures of these goals and concepts (the actual metrics to be used). The approach was to find reliable historical measures for these concepts and goals for the systems in Table 4-1, as far back as VAMOSC reported O&S cost data on them. To a great extent, data availability determined what metrics were used in this research. Therefore, a discussion about the sources of data at our disposal should serve as a template for discussing the specific metrics used in the analysis.

---

* Though system effectiveness was identified as one of the supra-ordinate goals of naval acquisitions, unfortunately, time did not allow for an analysis of this goal and it's associated subordinate metrics. However, since the data was collected, the metrics are included in this chapter, even though they are absent from the statistical analyses.

Data collection began in earnest upon obtaining Navy VAMOSC Shipboard Systems Reports (SSR) for all of the systems in Table 4-1. The VAMOSC SSR provide yearly O&S costs as prescribed by the OSD Cost Analysis Improvement Group (CAIG). The reports provide O&S costs broken down into several elements as specified by CAIG, in addition to some technical data. The following sample report shows the O&S costs as they appear in the SSR.

Table 4-2 – Sample VAMOSC REPORT

AN/BQQ-5 SONAR SYSTEM		
		1997
		Constant FY 00 Dollars
	AVERAGE SYSTEM COST	429,810
	TOTAL SYSTEM COST	21,920,325
1001	DIRECT SYSTEM OPERATION	11,801,712
1001.1	MANPOWER	11,801,712
1001.2	FOSSIL FUEL	N/A
1001.3	EXPENDABLE ORDNANCE	N/A
1002	DIRECT SYSTEM MAINTENANCE	8,890,552
1002.1	ORG REPAIR PARTS	594,861
1002.2	INTERMEDIATE MAINTENANCE	376,733
1002.2.1	AFLOAT IMA REPAIR PARTS	137
1002.2.2	ASHORE IMA REPAIR PARTS	49,356
1002.2.3	IMA LABOR	327,239
1002.3	DEPOT MAINTENANCE	2,480,921
1002.3.1	SYSTEM OVERHAUL AND REPAIR	0
1002.3.1.1	PUBLIC SHIPYARDS	0
1002.3.1.2	PRIVATE SHIPYARDS	0
1002.3.1.3	SHIP REPAIR FACILITIES	0
1002.3.2	FLEET MODERNIZATION	0
1002.3.2.1	ALTERATIONS	0
1002.3.2.1.1	PUBLIC SHIPYARDS (FMP)	0
1002.3.2.1.2	PRIVATE SHIPYARDS (FMP)	0
1002.3.2.1.3	SHIP REPAIR FACILITIES (FMP)	0
1002.3.2.2	CENTRALLY PROVIDED MATERIAL	0
1002.3.3	SYSTEM COMPONENT REWORK	1,981,823
1002.3.3.1	LOGCEN	1,568,741
1002.3.3.2	PM INPUT	413,083
1002.3.4	OTHER DEPOT MAINTENANCE	499,097
1002.4	ENGINEER AND TECH SERVICES	5,075,927
1002.5	EMBEDDED CPTR S/WARE SUPPORT	362,111

1003	REPLENISHMENT SPARES	161,544
1004	INDIRECT SYSTEM COSTS	1,066,517
1004.1	PCS	389,019
1004.2	TRAINING	677,498
2100	SHIPBOARD SYSTEM OPERATING DATA	
2101	NUMBER OF SYSTEMS	51
2104	TOTAL PERSONNEL ASSIGNED	295
2105	PERSONNEL TRAINED	52
2200	SHIPBOARD SYSTEM MAINT DATA	
2201	MAINTENANCE PHILOSOPHY	
2202	ORG CORRECTIVE MAINT MANHOURS	40,910
2203	IMA MAINT MANHOURS	23,869
2203.1	AFLOAT MAINTENANCE MANHOURS	1,644
2203.2	ASHORE MAINTENANCE MANHOURS	22,225
2205	NUMBER OF CORRECTIVE MAINT ACTIONS	2,056

The data in the SSR are gathered from several sources throughout the Navy, including:

- Bureau of Naval Personnel (BUPERS)
- Chief of Naval Personnel (PCSVC)
- Defense Finance and Accounting Service-Cleveland Center (DFAS-CL)
- Defense Finance and Accounting Service Navy Cost Information System-Operations Subsystem (NCIS-OPS)
- Naval Education and Training Professional Development Technology Center (NETPDTC)
- Naval Ordnance Center IMSD, Mechanicsburg
- Naval Surface Warfare Center (NSWC) Carderock Div Philadelphia
- Naval Sea Systems Command (SEA 04)
- Naval Sea Systems Command Logistics Center (NAVSEALOGCEN)
- Naval Shipyards
- Naval Surface Warfare Center (NSWC) Port Hueneme Detachment (PHD) Louisville
- Planning and Estimating for Repairs and Alterations (PERA)
- Program Offices
- Supervisors of Shipbuilding, Conversion and Repair (SUPSHIPS)

In addition to the information in the SSR, the Navy VAMOSC program provided other information it had gathered pertaining to the number of personnel required to maintain and operate each system (according to the program office), the amount of training given to those who operate and maintain the systems, and also the weight of some systems (a surrogate measure of the system's complexity). For more information about Navy VAMOSC SSR, refer to the VAMOSC "Data Reference Manual for Shipboard Systems" available at www.ncca.navy.mil/vamosc.

The Navy's Combat Systems Troubled Systems Process (TSP) provided the second major source of data for this research. The mission of TSP is "To improve Fleet readiness by identifying Navy-wide combat system maintenance problems, their root causes; and in conjunction with the Top Management Attention (TMA)/Top Management Issues (TMI) program, develop and monitor the Plan of Action and Milestones (POA&M) to correct those problems." TSP accomplishes this with the metrics it collects on Navy systems. TSP uses the information to identify so-called "bad actors," systems whose bad performance (as measured by the TSP metrics) requires remedial action. In addition, TSP also tracks the root causes of maintenance problems. TSP uses the information gathered during shipboard inspections of equipment, called Combat Systems Readiness Reviews (CSRR), as its primary source. In addition, TSP pools data from the Navy's Board of Inspection and Survey (INSURV), Worldwide Technical Assist (WWTA) database, and Casualty Reports (CASREPS).

The Navy's Material Readiness Data Base (MRDB) served as the third major source of data for this research. The MRDB compiles corrective maintenance data from the fleet, scrubs the data for errors, and uses it to compute reliability, maintainability, and supportability readiness indices. The data the MRDB uses to calculate these readiness measures are subjected to rigorous examination to verify that they are correct before they are used. Although the rigorous screening of data that the MRDB performs increases the reliability of the MRDB metrics, the amount of time and funding required to do this limits the number of systems for which the data is available.

The three major data sources described thus far all pertain to cost, reliability, maintainability, availability, and supportability. These sources do not provide information about some of the other, less easily measured cost and effectiveness drivers identified in our discussions with Navy representatives. They offer very little information on important, but hard to measure priorities like:

- o Capability of Performance
- o The "People Factor"
- o Ease of Technology Insertion/Refreshment, or "Upgradability"
- o Openness of System Architecture
- o Inclusion of COTS/NDI Technology
- o Software Support and Maintenance Requirements.

Although we found no existing metrics or sources for this kind of data, we gathered from our discussions with various programs in the Navy that these concepts were too important to exclude from our statistical model. We therefore turned to survey questions to measure the degree to which the systems in Table 4-1 had exploited these concepts.

As suggested earlier, the definitions and meanings of these terms varied somewhat among the individuals and organizations we met with. Thus, the subjectivity inherent to assessing these concepts necessitated consistent, objective criteria by which to measure them. To determine the appropriate measurement criteria, representatives from the Open Systems – Joint Task Force, the Affordability Through Commonality Program, the Naval Post Graduate School, as well as several Navy program managers and ISEA's were consulted. We asked each group to suggest criteria for measuring the above concepts on a 5-point scale. Specifically, we asked them to

37

suggest criteria to associate with the end points and/or mid point of a 5 point scale according to realistic extremes and/or average values one would observe in equipment that is currently used by the Navy. For example, representatives from the Open Systems – Joint Task Force helped generate the following question to assess the openness of a system's architecture:

Use of Open Architecture and Open Standards

Description: The degree to which open architecture/open standards were used in the design in order to allow easier upgrades and multiple suppliers of hardware and software (e.g. use of VME or VXI standards or use of an RS232 interface). Please evaluate (1 to 5) on according to the following scale:

Scale	Description
1	Exclusive use of proprietary or system unique hardware and software interfaces and standards.
2	Extensive use of proprietary or system unique hardware and/or software interfaces and standards.
3	Limited use of proprietary or system unique hardware and/or software interfaces. Some degree of Open architecture open standards used.
4	Open architecture open standards used significantly. Minimal use of proprietary or system unique hardware and/or software interfaces.
5	Extensive use of open architecture, open standards. System architecture allows for continuous upgrades throughout life cycle and support by multiple suppliers.

To further reduce the subjectivity of these measures, the survey questions were administered (whenever possible and appropriate) to three different individuals from the organizations that support and manage each of the systems in Table 4-1; one from the program office, one from the ISEA, and one "waterfront" technician from FTSCLANT. Moreover, obtaining three different responses to each question (whenever possible and appropriate) made possible a statistical assessment of the reliability of the questions (discussed in Chapter 5).

This section lists each of the metrics that were gathered as data for this research as well as each metric's source and a brief description of what it actually measures. Each metric is supposed to measure (at least partially) the degree to which a system exhibits some theoretical construct, or concept. (Note that for the purposes of this research, the words concept and construct are used interchangeably and both signify an attribute, priority, or goal to be measured.) For example, the metric mean-time-between-failures (MTBF) is used to measure the degree to which the system is reliable. In some cases, multiple metrics are necessary to capture the different aspects of a construct. In other cases, a metric may (partially) measure more than one construct. For a more logical presentation, metrics that measure similar constructs or different aspects of the same construct are grouped together in this list under the appropriate construct (exactly how well the metrics agree with each other within these groups is discussed in Chapter 5). The reader will quickly realize that these metrics are imperfect measures of the underlying constructs they intend to measure. Often, the most obvious measure for a construct simply isn't available, so surrogate measures are used. In many cases, the metrics intended to measure the degree to which a construct is inherent to a system may actually reflect Navy policy as much as any intrinsic attribute of the system. Returning to the MTBF example, MTBF may reflect the conditions under which the system is operated (i.e. accomplishment of preventative maintenance, how often the system is used, etc., etc.) as much as it measures the system's inherent reliability. The metrics gathered for this research represent the results of months of searching for the best available metrics to measure the constructs that drive system cost-effectiveness.

4-4-1 Supra-Ordinate Constructs and Metrics

Construct: Total Ownership Cost
Metric: Yearly Operating and Support Cost per System
Source: Navy VAMOSC Shipboard System Reports (SSR's)
Description: The total (Navy-wide) yearly O&S cost of a particular system (e.g. BQQ-5 sonar, SPS-55 radar, or MK 45 Gun) divided by the number of systems in service. The O&S cost reported by VAMOSC includes the cost of manpower and training for personnel assigned to the system, direct maintenance costs (parts and labor at all maintenance levels), expendable ordnance costs (when applicable), modernization expenses, and computer, software and technical support costs. Since most systems in the VAMOSC reports do not launch any kind of missiles or munitions, I subtracted out the ordinance costs for the few systems that do, in order to make a more reasonable comparison with systems that do not fire any kind of munitions.

Construct: System Mission Effectiveness
Discussion: As time did not permit an analysis of system effectiveness metrics, the effectiveness metrics in this chapter were not used in the data analysis. Nonetheless, effectiveness metrics are reported in this chapter for use in future analysis.

As mentioned in Section 4.2, the Port Hueneme Division of the Naval Surface Warfare Center (PHD NSWC) proposes three aspects of system effectiveness in its "Safety, Effectiveness and Affordability Review Guide;" the inherent Performance Capability (P_c) of the system, the "People Factor" (P_f), and the system's Operational Availability (A_o). The effectiveness metrics

used in this research were intended to measure these three aspects of effectiveness. Of these three effectiveness constructs, only A_o is widely tracked in the Navy, but even so, it is only available for 16 of the systems in the VAMOSC SSR's. Therefore, it was necessary to adapt the first two measures, P_c and P_f, to a questionnaire. In addition, a fifth measure of effectiveness, Equipment Operational Capability (EOC) was also used as a supplement to A_o. This measure, provided by the Navy Troubled Systems Process (TSP) is similar to A_o, but not exactly the same.

Three questions were asked of representatives from each of the three agencies responsible for managing the system (the system's program office, the in-service engineering agent (ISEA), and the support technicians who maintain the system). Two of these questions asked the respondents to rate the system's Performance Capability and People Factor, based on the criteria suggested by the "PHD NSWC Safety, Effectiveness, and Affordability Review Guide" and by engineers from PHD NSWC. The third question asked the respondents to rate the over-all effectiveness of the system according to all three aspects of effectiveness (P_c, P_f, and A_o). The following are the metrics/questions used to assess system effectiveness:

Metric 1: System Inherent Performance Capability
Source: Questionnaire administered to system program managers, ISEA's and support technicians.
Description: Question designed to measure the Performance Capability of the system using the criteria suggested by NSWC PHD. Respondents were given the following description and scale (respondents were told they could rate their system anywhere on the 1-5 scale even though some scales did not have criteria for values of 2 and 4):

> A qualitative measure of how well the system performs the mission it was designed for when it is fully functional. *Assuming the system is fully functional*, please rate its performance according the following index:

Index Value	Description
1	Marginal effectiveness in performing its given mission, much room left for improvement. There is considerable risk that the system will not perform its mission well.
2	
3	Satisfactory effectiveness in performing given mission, some room left for improvement. Still some uncertainty as to whether the system will perform its mission well.
4	
5	Excellent effectiveness in performing given mission. Consistently performs well at the mission for which it was designed.

Metric 2: The People Factor (or "Sailor Proofness")
Source: Questionnaire administered to system Program managers and ISEA's and Support Technicians.

Description: Question designed to measure the People Factor performance area of the system using the criteria suggested by engineers from NSWC PHD. Part of what NSWC PHD calls the People Factor, is commonly known in the Navy as the "Sailor-Proofness" of a system. Respondents were given the following description and scale:

The degree to which the system is immune to failures due to sailor error and/or inexperience. Please rate the system according to the following criteria:

Index Value	Description
1	Not sailor proof at all. System relies heavily on reminders/warnings/placards and/or assumes the operator has a lot of special knowledge or familiarity with the system to operate and maintain the system.
2	
3	
4	
5	Very sailor proof. System requires minimal use of "work arounds," or special placarding. System does not require excessive knowledge of the system to operate and maintain it.

Metric 3: Operational Availability (A_o)
Source: The Material Readiness Data Base (MRDB)
Description: Measures the likelihood that a system will be available for use, at a pre-determined level of performance and for a sufficient length of time, whenever it is needed.

MRDB computes A_o using the following quantity: $\dfrac{MTBF}{MTBF + MDT}$ where MDT is mean-down-time, the average downtime for the system after system failure.

Metric 4: Equipment Operational Capability (EOC)
Source: Navy TSP
Description: EOC is a number assigned to equipment during the Navy's CSRR inspections. EOC's are assigned to subsystems and then used to calculate an EOC at the shipboard system level. At the system level, EOC ranges from 0 to 1 and measures a system's ability to pass pre-defined tests that indicate whether or not the system can provide a certain level of functional capability. An EOC of 0 signifies that a system has no functional capability at all, while an EOC of 1 signifies that a system is fully operational. EOC is somewhat different than A_o in that EOC provides a "snapshot" indication of a system's capacity to function at the particular time it was inspected. On the other hand, Ao measures the percentage of time the system was available over a period of time (typically a fiscal or calendar year). To an extent, both metrics measure a system's ability to operate at a pre-defined level of performance. EOC has the advantage of

being more widely available (EOC is available for almost all of the systems in the VAMOSC SSR's).

Metric 5: Over-All System Mission Effectiveness
Source: Questionnaire administered to system program managers, ISEA's and support technicians.
Description: Respondents were asked to rate their system's effectiveness according to all three aspects of effectiveness, using the following description and scale:

Please rate the system's mission effectiveness on the following scale (in the mission it was designed for) considering the system's availability (Ao), mission performance when it is available, and the probability that the operators will perform correctly:

Index Value	Description
1	Marginal effectiveness in performing its given mission, much room left for improvement. There is considerable risk that the system will not perform its mission well due to availability problems, operator error, or inadequate performance capability.
2	
3	Satisfactory effectiveness in performing given mission, some room left for improvement. Still some uncertainty as to whether the system will perform its mission well due to availability problems, operator error, or inadequate performance capability.
4	
5	Excellent effectiveness in performing given mission. Little to no doubt the system will perform well when called upon. High availability, and performance capability.

4-4-2 Subordinate Constructs and Metrics

Construct: System Manpower Requirements
Discussion: Many of the individuals consulted at NAVSEA identified manpower as the single biggest driver of TOC. Therefore, some measures of a system's manpower requirements were necessary for cost modeling. A natural metric for measuring a system's manpower requirements is the number of personnel a system requires for its operation and maintenance. One might expect to measure this by simply asking the individuals who manage the system "How many personnel does it take to operate and maintain this system?" In practice, however, there is not usually a straight forward answer to this question (at least for many of the systems in the study). Many systems do not have a person officially assigned to them and are operated and maintained by personnel who divide their time among several systems, in addition to other duties on the ship (such as re-painting the ship or working in the mess hall). Some systems only require human

operators during one particular mode of operation. For example, the SLQ-25 system does not require any personnel to operate it, except when it deploys, during which time it requires 5 personnel to operate. In some cases, manpower requirements may even depend on what class of ship on which the system is operating. Therefore, measuring a system's exact manpower requirements is a difficult proposition.

Two alternative metrics were available for measuring system manpower requirements, both supplied by VAMOSC. The first metric is the number of personnel officially assigned to each system and the second is the response of the program office to the question, "How many personnel does it take to operate in maintain this system?" Exactly what each metric measures and the respective pros and cons of the two metrics are discussed below.

Metric 1: Personnel Assigned NEC's per System
Source: VAMOSC supplies the total number of personnel assigned to a system in its SSR's.
Description: Personnel Assigned per System is simply the total number of personnel assigned to a system divided by the total number of systems Navy wide (as reported in the SSR). VAMOSC obtains the total number of personnel assigned to a system from the Bureau of Naval Personnel (BUPERS). BUPRERS provides the total number of personnel who have a Naval Enlisted Classification code (NEC) corresponding to the system. This metric actually measures the number of people who have a NEC for the system. A NEC indicates that a person has had training to operate and/or maintain a system. It does not imply that the person devotes all his/her time operating and/or maintaining that particular system. The same person may perform tasks on other systems. Some smaller systems do not even have NEC's (though not necessarily the systems in this study), yet still require manpower for operation and maintenance. Thus, Personnel Assigned per System is an imperfect measure of a system's true manpower requirements. The advantage to this metric is that it is somewhat less subjective than asking the program office the ambiguous question, "How many personnel does this system require for maintenance and operation?"

Metric 2: Personnel per System According to the Program Office
Source: VAMOSC initially furnished this data with its SSR's but has omitted it from later SSR's. VAMOSC furnished this data upon my request. VAMOSC obtained the data by asking the Program Office how many personnel are required to operate and maintain the system as stipulated by the Navy Training Plan.
Description: As previously mentioned, this too is an imperfect measure of a system's manpower requirements. This number may depend on what mode of operation the system is in, the particular variant of the system, or even the ship the system is installed on. In many cases, the number provided was a range, for example, 8-10 personnel to operate and maintain each system. In a few cases, the range was even wider. In each case in which a range was provided, I took the midpoint of the specified range as the value of this measure for the system in question. While this metric is obviously imprecise, it has the merit of being closer (at least theoretically) to the underlying construct than the Number of Personnel Assigned per System metric. Therefore, both metrics were included in this research as measures of system manpower requirements.

Construct: System Training Requirements
Discussion: A system's training requirements relate directly to the People Factor measure of effectiveness as well as the system's O&S cost. In much the same way that manpower requirements are difficult to measure, the same is true of system training requirements. VAMOSC provided the two measures described below.

Metric 1: Training Courses Completed per System
Source: VAMOSC SSR (obtained by VAMOSC from the Naval Education and Training Professional Development Technology Center (NETPDTC))
Description: The following excerpt from the VAMOSC SSR Data Reference manual describes this metric:

> The number of personnel trained as reported by NETPDTC reflects graduates from the appropriate courses related to the shipboard system. An individual sailor may graduate from more than one course and will be counted as a graduate from each course. Also a ship may send individuals for team training several times during a fiscal year. Each individual member of the team may be counted as a graduate each time the team completes a course.
>
> Therefore, the number displayed in this element does not indicate the number of individual sailors that were trained in the shipboard system related courses.

This metric is at best, a surrogate for measuring the amount of training per system (though it reveals very little about the depth of the training). It does, however, have the advantage of being readily available as far back as the VAMOSC SSR's are available.

Metric 2: Student Training Days per System
Source: VAMOSC (not available in SSR) (obtained by VAMOSC from NETPDTC)
Description: This metric measures the total amount of student-training-days per system in each year. It is the sum of all the days of training received by all personnel who received training for a particular system. It is somewhat closer (at least in theory) to measuring the amount of training received by the maintainers and operators of the system, but unfortunately, it was only available for the years 1995-1999.

Construct: The Degree to Which the System is Automated
Discussion: This concept is a potential avenue to reduce manpower requirements, and thus, cost. In addition to its potential to reduce manpower costs, making a system more automated in its operation will also reduce the number of personnel who are exposed to danger in a wartime scenario.
Metric: Degree of Automation
Source: Questionnaire administered to program managers and ISEA's
Description: A qualitative measure to capture whether a system is automated or manpower intensive in its operation. Respondents were asked to assess their system according to the following description and scale:

Please rate the degree to which the system is automated in its operation. Consider the manpower required to operate the system.

Index Value	Description
1	Manpower intensive. System requires human operators to perform many tasks that could be performed automatically.
2	
3	Some automation in the design. Some visible effort in the design to reduce the manpower required to operate the system.
4	
5	Highly automated. Design has many features to reduce manpower required to operate the system. System requires very few tasks to be performed by humans that could be performed automatically.

Construct: System Software Support and Maintenance Requirements
Discussion: One of the cost elements in the VAMOSC SSR's is Embedded Computer Software Support. In perusing the SSR's for the systems in Table 4.1, it became apparent that the systems with the highest O&S costs typically exhibited very large expenditures in this cost element. Furthermore, within the SSR of a particular system, the same held true; the years during which the system's cost were highest often corresponded to large expenditures in this area. This was especially true of large sonar systems, which as a category, tended to incur the highest O&S costs on a per system basis. Subsequent discussions with program managers and ISEA's validated that the software support requirements of a system played an important role in its O&S cost. Discussions with software experts at NAVSEA initially directed me to potential measures of a system's software development and support requirements. Both the size of the system's software and the complexity of the system's software were proposed as appropriate metrics. Additionally, the robustness of the software's code to changes in hardware was also deemed important. Visits to the websites of the Carnegie Mellon Software Engineering Institute, the Texas A&M Department of Computer Science, and the Softstar Systems' COCOMO software cost estimating model provided valuable suggestions for metrics to measure the size and complexity of a system's software.

Metric 1: Thousands of Lines of Source Code (KSLOC)
Source: Provided by program managers and/or ISEA's.
Description: An estimate of the size of the system's software.

Metric 2: Software Complexity
Source: Questionnaire administered to system program managers and ISEA's
Description: While there exist very precise measures for assessing the complexity of software (such as the number of modules, the arc-to-node ratio, etc., etc.), it was not practical to gather this information on 50 or so systems, and asking this information from the program managers

and ISEA's would have generated an unreasonable amount of work for them (moreover, in many cases, only the companies that designed the system would have this information, making it even more difficult to ascertain). In lieu of burdening the program managers and ISEA's with this type of request, they were asked to rate the complexity of their system's software using the following description and scale (developed with the assistance of NAVSEA ISEA):

Please rate the complexity of the system's software according to the following scale:

Index Value	Description
1	Very simple. Most code passes data back and forth, very few algorithms.
2	
3	
4	
5	Very complex. Code is comprised primarily of sophisticated algorithms that perform complex computations.

Metric 3: Software Robustness to Hardware Obsolescence
Source: Questionnaire administered to system program managers and ISEA's
Description: Changes in system hardware that require rewriting of code are among the largest drivers of software support costs. To an extent, the cost of rewriting software can be minimized by code that is robust to hardware changes and reusable. To quantify how the extent to which the system's software was robust to hardware obsolescence, the program mangers and ISEA's were asked to rate their system according to the following description and scale:

Please rate the degree to which the system's software has been robust to changes in hardware. Consider how much of the system's software has been rewritten with changes in the system's hardware.

Index Value	Description
1	Not robust at all. Substantial writing/rewriting of code required each time the hardware changes.
2	
3	
4	
5	Very robust. Minimal rewriting of code due has accompanied hardware changes in the past.

Construct: System Corrective Maintenance
Discussion: Much of the data available from the Navy pertains to how many corrective maintenance actions a system requires and how many hours are spent performing them. The corrective maintenance that a system requires contributes directly to the system's cost (repair parts and labor) and also indirectly, by driving manpower requirements. Each system also generates a certain amount of preventative maintenance, however, the actual amount of preventative maintenance performed on a system is very difficult to measure. In theory, each system has a certain amount of required preventative maintenance actions and an estimated amount of time necessary for performing them. In practice, however, preventative maintenance is not always accomplished (at least, according to several of the program managers and technicians interviewed) and it would be almost impossible to know exactly how much is actually performed on each system. Therefore, most of the data pertains to corrective maintenance as opposed to preventative maintenance.

Metric 1: Corrective Maintenance Actions per System
Source: VAMOSC SSR (obtained by VAMOSC from the Naval Sea Logistics Center Ship's Ships' 3M Data Base)
Description: The number of corrective maintenance actions performed on a system divided by the number of systems.

Metric 2: Corrective Maintenance Man-Hours per System
Source: VAMOSC SSR (obtained by VAMOSC from the Naval Sea Logistics Center Ship's Ships' 3M Data Base)
Description: The number of corrective maintenance man-hours performed per system (at the oganizational level).

Metric 3: Casualty Reports per System (CASREPS)
Source: Navy Combat Systems Troubled Systems Process (TSP) (TSP obtains this from the Naval Sea Logistics Center)
Description: When a system fails and as a result, the ship's mission capability is compromised, the Captain may elect to make a Casualty Report if in his judgment, the problem is grave enough to warrant one. Thus, the number of CASREPS per system is a subset of the number of Corrective Maintenance Actions per System. CASREPS represent corrective maintenance actions urgent enough to require immediate attention. Several different organizations monitor the number of CASREPS per system to measure the system's performance. However, CASREPS do not directly measure the amount of maintenance actions a system generates since only those that degrade the ship's mission capability are reported (and this depends heavily on the Captain and the mission the ship happens to be performing at the time of the failure).

Metric 4: CASREP Maintenance Man-Hours per System
Source: Navy TSP (TSP obtains this from the Naval Sea Logistics Center)
Description: The total number of maintenance man-hours generated by the system's CASREPS divided by the number of systems. Since only those failures grave enough to impede the ship's mission capability are reported as CASREPS, this may not represent the system's total corrective maintenance man-hours as well as the man-hours supplied by the VAMOSC SSR's.

47

Metric 5: Maintenance Workload
Source: Questionnaire administered to system program managers and ISEA's and support technicians.
Description: Respondents were asked to assess the maintenance workload generated by their system according to the following description and scale:

A qualitative measure to capture the system's impact on sailor workload from the maintenance required to keep it operating. Please rate the system's maintenance workload on the following scale:

Index Value	Description
1	Manpower intensive. System requires many man-hours of preventive and corrective maintenance.
2	
3	Moderate maintenance manpower impact.
4	
5	Very low maintenance. System requires very few man-hours of preventive and corrective maintenance.

Construct: Inherent Reliability of the System
Discussion: The reliability of the system plays a critical role in determining how often the system fails (along with the conditions and tempo under which the system is operated), and therefore, the system's cost and effectiveness. Therefore, some measure of system reliability should factor into an analysis of cost-effectiveness.
Metric: Mean-Time-Between-Failures (MTBF)
Source: MRDB
Description: Though MTBF is the most common measure of system reliability, it is not an exact measure of the system's inherent reliability since the conditions under which the system operates also determine the system's MTBF. Moreover, MTBF registers only system failures and not lesser failure events that require corrective maintenance but do not cause the system to fail. Unfortunately, MTBF (and all other MRDB data) was only available for 16 of the systems in Table 4.1.

Construct: Inherent Maintainability of the System
Discussion: In addition to how reliable a system is, it is important that the system's maintenance be simple enough that the ship's crew is able to accomplish it in a timely manner, without relying on outside assistance. The more complicated the system's maintenance is, the more time it will take to repair, the more training it will require of the crew, the more man-hours it will require to accomplish, the more outside assistance the system will require, and therefore, the more the system will cost to operate and maintain. It should be noted, however, that most of the available metrics for measuring the inherent maintainability of a system are also driven by the quality and

48

quantity of training, and may, therefore, reflect Navy policy as much as the system's inherent maintainability.

Metric 1: Mean-Time-To-Repair (MTTR)
Source: MRDB
Description: The mean time to repair the system for corrective maintenance actions.

Metric 2: Corrective Maintenance Man-Hours per Corrective Maintenance Action
Source: VAMOSC SSR
Description: The total number of corrective maintenance man-hours divided by the total number of corrective maintenance actions reported in the VAMOSC SSR. This metric is intended to measure the same thing as the MTTR metric furnished by the MRDB. While this data has not been subject to the same scrutiny as the MRDB data (and is probably not as accurate as the MRDB), it has the benefit of being available for all the systems since it came in the SSR's. Therefore, it was used as a supplement to the MRDB MTTR data.

Metric 3: CASREP Maintenance Man-Hours per CASREP
Source: Navy TSP (TSP obtains this from the Naval Sea Logistics Center)
Description: The total number of CASREP maintenance man-hours divided by the total number of CASREPS reported by Navy TSP.

Metric 4: Technical Assist Visit Requests (TAVR) per System
Source: Navy TSP
Description: A TAVR occurs when ship personnel require outside assistance to repair a system and it takes the assisting technician 4 hours or more to finish the repair. This metric is an indirect measure of the system's maintainability, but may also reflect the amount of training crews are receiving to perform the maintenance. Moreover, TAVR/System is an indirect measure of reliability in that in order for a TAVR to occur, a failure must first occur.

Metric 5: Maintenance/Operator Induced Failures per CSRR
Source: Navy TSP
Description: As part of the CSRR inspections the Navy performs on its ships and systems, the technicians who perform the inspections document the root causes of all equipment deficiencies. One of the root cause categories is "Maintenance/Operator Induced Failures." This metric records the number of times maintenance or operator induced failures were cited as the root cause of deficiencies found in CSRR inspections, normalized by the number of inspections performed on that system (Navy wide). Navy analysts familiar with the TSP cautioned that this number may be underreported as inspectors may be reluctant to attribute equipment problems to the crew.

Construct: Effectiveness of Built-In Testing (BIT)

Discussion: Increasing the effectiveness of the system's BIT will, in theory, make the system's maintenance easier to perform.

Metric 1: BIT Quality

Source: Questionnaire administered to system program managers and ISEA's, and support technicians.

Description: Respondents were asked to assess the quality of their system's BIT according to the following description and scale:

> Please rate the quality of the BIT in terms of the typical size of the ambiguity group when the system fails. Please rate the typical size of the ambiguity group in actual operation as opposed to what the specification requires.

Index Value	Description
1	Poor BIT. (e.g. when a fault occurs, the ambiguity group can be 10 LRU's or more.)
2	
3	Medium quality BIT. (e.g. ambiguity groups are typically about 5 or 6 LRU's.)
4	
5	Excellent BIT. (e.g. ambiguity groups are 3 or less LRU's, 95% of the time or more.)

Metric 2: Number of Inadequate BIT Problems Reported per Combat System Readiness Review (CSRR)

Source: Navy TSP

Description: This metric records the number of times inadequate BIT equipment problems were reported by inspectors, normalized by the number of inspections performed on that system (Navy wide).

Construct: Modularity
Discussion: Modularity is an attribute that (in theory) facilitates maintenance and also allows for easier upgrades and modifications to the system (or parts of the system).
Metric: Degree of Modularity
Source: Questionnaire administered to system Program managers and ISEA's and Support Technicians. Respondents were asked to assess the modularity of their system according to the following description and scale:

Description: The degree to which the system is modular. Please evaluate according to the following index:

Index Value	Description
1	Very little modularity to the design. Functionality widely distributed throughout different parts of the system. Failure of one of the system's functions can require working on several different parts of the system.
2	
3	Design is partially modular. Some degree of isolation of functionality in different modules.
4	
5	Design is very modular. Different functions performed by distinct, different modules that can be easily removed or replaced.

Construct: Ease of System Upgrade/Technology Insertion
Discussion: As system life cycles have lengthened (spanning decades in many cases), and the cycle time for the development of new technology (especially information technology) has decreased, it has become increasingly important that a system be easily and inexpensively upgraded as new threats and technologies emerge. The concept of designing a system for affordable upgrades has become a priority of defense acquisition reform in recent years. There are many factors that determine the ease to which a system can be modified and upgraded, however, there were no pre-existing data available at the outset of this research. Therefore, the following questions were developed with the help of representatives from the DoD Open Systems – Joint Task Force, the Navy's Affordability Through Commonality Program, and other Navy engineers and program managers.

Metric 1: "Upgradability" or Ease of Technical Refreshment
Source: Questionnaire administered to system program managers and ISEA's
Description: Respondents were asked to assess the degree to which their system's design lends itself to easy and inexpensive upgrade and modification according to the following description and scale:

The degree to which the system's architecture allows for easy integration of new technology to improve performance and reliability. Please evaluate according to the following scale:

Index Value	Description
1	Very difficult to upgrade. Incorporating new technology requires replacement of the entire system (e.g., replacing a processor would require major changes to the system software).
2	Difficult to upgrade. Incorporating new technology requires substantial revamping of the system. Changes to one part of the system require many changes to the rest of the system.
3	Some, but limited upgradability without completely changing the system. Possibility of upgrading part of the system without having to change the rest.
4	Incorporating new technology and later improvements in reliability and performance was considered in the design.
5	Easy to upgrade, preplanned product improvement (P^3I) designed into the system. Incorporating future technology was a priority in the design.

Metric 2: Use of Open Architecture and Open Standards
Source: Questionnaire administered to system program managers and ISEA's
Description: Respondents were asked to assess the "openness" of their system according to the following description and scale:

The degree to which open architecture/open standards were used in the design in order to allow easier upgrades and multiple suppliers of hardware and software (e.g. use of VME or VXI standards or use of an RS232 interface). Please evaluate according to the following scale:

Index Value	Description
1	Exclusive use of proprietary or system unique hardware and software interfaces and standards.
2	Extensive use of proprietary or system unique hardware and/or software interfaces and standards.
3	Limited use of proprietary or system unique hardware and/or software interfaces. Some degree of open architecture/open standards used.
4	Open architecture open standards used significantly. Minimal use of proprietary or system unique hardware and/or software interfaces.
5	Extensive use of open architecture, open standards. System architecture allows for continuous upgrades throughout life cycle and support by multiple suppliers.

Metric 3: Use of Commercial Components and Parts
Source: Questionnaire administered to system program managers and ISEA's.
Description: Respondents were asked to assess the degree to which their system uses commercial parts and components according to the following description and scale:

Please rate the system on the following 1 to 5 scale according to the extent to which it uses commercial components and parts. Please use the following definition of commercial parts in your answer:

"Commercial component" means any item "offered for sale, lease, or license to the general public" or any item that will soon be available in the commercial marketplace including commercial items requiring minor modifications to meet Federal Government requirements.

Index Value	Description
1	No usage of commercial components/parts or software. All of the system was designed using Milspec parts and components and software.
2	Very Low usage of commercial components/parts or software (e.g. some COTS power supplies or a COTS computer monitor). Most, but not all, of the system's parts, components, and software was designed from scratch, specifically for use in this system.
3	Moderate use of commercial components/parts or software. For example, one or more cabinets composed largely of commercially available components.
4	Substantial use of commercial items. Many components or parts or software modules of this system were developed or are in use commercially.
5	COTS with very minor modification.

Metric 4: Construct: Use of Non Development Items (NDI)
Source: Questionnaire administered to system program managers and ISEA's
Description: Respondents were asked to assess the degree to which their system uses NDI
according to the following description and scale:

> Please rate the system on the following 1 to 5 scale according to the extent to which it
> uses components and parts that were not designed specifically for this system, but are in
> use in other systems or are commercially available.

Index Value	Description
1	No usage of NDI. System was designed completely from scratch using parts, components, and software designed specifically for use in this system.
2	Very Low usage of NDI. Most, but not all, of the system's parts, components, and software was designed from scratch, specifically for use in this system.
3	Moderate usage of NDI. Some subsystems, parts, and/or software used in the system were or are in use in other systems, whether commercial or other Navy systems (e.g. system uses a computer or other minor subsystem developed elsewhere).
4	Significant usage of NDI. System uses a significant amount of parts, components, and/or software that are NDI (e.g. a major subsystem was developed elsewhere or many subsystems were NDI).
5	Substantial use of NDI. Many major components or parts of this system were developed or are in use elsewhere, whether in other Navy systems or commercially.

Construct: Usability
Discussion: In the data collected, there were some data that (may) indicate how well humans operators and maintainers interface with the systems. These were grouped under the heading, "usability."

Metric 1: Sailor Proofness

Metric 2: Maintenance/Operator Induced Failures per CSRR

Metric 3: Inexperienced Personnel Problems per CSRR
Source: Navy TSP
Description: The number of times inexperienced personnel were cited as the root cause of deficiencies found in CSRR inspections, normalized by the number of inspections performed on that system (Navy wide). As with the number of maintenance/operator induced failures, Navy

analysts familiar with the TSP cautioned that this number may be underreported as inspectors may be reluctant to attribute equipment problems to the crew.

Metric 4: Inadequate Training Problems Reported per CSRR
Source: Navy TSP
Description: The number of times inadequate training was cited as the root cause of deficiencies found in CSRR inspections, normalized by the number of inspections performed on that system (Navy wide). Like the number of inadequate manning problems per inspection, this metric may indicate shortfalls in Navy training (or perhaps funding), but may also be interpreted as a reflection of high training requirements (or a system that is not easy to maintain and/or operate) as one would expect inadequate training problems to be more prevalent among systems with high training requirements.

Construct: Supportability
Discussion: The ability of the Navy to supply spare parts when needed is critical to system operational availability, and therefore, effectiveness. As previously mentioned, time did not allow for the analysis of these and other effectiveness metrics, however, they are reported for further research.

Metric 1: Mean-Logistics-Delay-Time (MLDT)
Source: MRDB
Description: MLDT measures the average delay caused by waiting for a part. Like all other MRDB data, this metric is only available for 16 of the VAMOSC SSR systems.

Metric 2: Mean Logistics Time (MLT)
Source: MRDB
Description: MLT measures the average delay time per failure event in which a part must be requisitioned from the Navy supply system because it is not on board the ship. MLT differs from MLDT in that only those failures in which a part must be requisitioned are in the denominator of MLT, whereas all failures are included in the denominator of MLDT.

Metric 3: Percent Not on Board (%NOB)
Source: MRDB
Description: The percentage of repair parts used that were not already on board the ship and had to be requisitioned from supply.

Metric 4: Logistics Delay Ratio (LDR)
Source: MRDB
Description: The ratio MLDT/MLT. This equates to the percentage of system failures in which a delay results because a part must be requisitioned.

Metric 5: Supply Hours per CASREP
Source: TSP
Description: The number of supply delay hours per CASREP.

Metric 6: Spare not Allowed Problems per CSRR
Source: Navy TSP
Description: Each ship carries an inventory of spare parts and each system has a list of allowed parts in the inventory. Since the ship does not carry a spare of every kind of part, a part is either "allowed" if it is carried by the ship or "not allowed" if it is not carried by the ship. This metric records the number of times inspectors attributed a CSRR deficiency to the lack of a replacement part that was not allowed on the system's spare part list.

Metric 7: Allowed Spare not on Board
Source: Navy TSP
Description: This metric records the number of times inspectors attributed a CSRR deficiency to the lack of a replacement part that was allowed on the system's spare part list, but not in the inventory.

4-4-3 Interrelationships Among the Metrics

Thus far, we have enumerated all the metrics gathered for this research. We have already begun to group similar metrics together, and we have already distinguished between supra-ordinate metrics and subordinate metrics that drive the supra-ordinate metrics. Thus, we have implicitly begun to categorize the relationships among metrics as either complementary or causal. Metrics that measure the same construct (or different aspects of the same construct) have a complementary relationship. For example, the metrics Personnel Assigned NEC's per System and Personnel per System According to the Program Office are complementary measures of a system's manpower requirements. On the other hand, there are also cause-and-effect relationships among the metrics. For example, one might hypothesize that System Corrective Maintenance may drive System Manpower Requirements. Under this hypothesis, the metric Corrective Maintenance Actions per System would have a causal relationship with the metric Personnel Assigned NEC's per System. The nature of the relationship between metrics is not always clear. For example, the metric Corrective Maintenance Actions per System measures the amount of maintenance actions a system generates and the metric Corrective Maintenance Man-Hours per System measures the amount of time those actions required. Both metrics reflect complementary aspects of System Corrective Maintenance. However, one might also argue that the metric Corrective Maintenance Actions per System drives the metric Corrective Maintenance Man-Hours per System, at least in part. Nonetheless, the two categories provided a useful framework in forming an initial mental model of how the metrics interrelate to each other and their relationships to the supra-ordinate metric of O&S cost.

The causal diagram in Figure 4-1 illustrates the complementary and causal relationships among metrics (as initially hypothesized). Those metrics that are complementary are grouped together (in boxes) under the construct they are supposed to measure. These complementary metrics are standardized and then summed, yielding a "purified" metric for the corresponding construct (this is discussed in greater detail in the following chapter). The arrows in the diagram represent (hypothetical) cause-and-effect relationships among the constructs (and the associated metrics). The diagram reveals a hierarchy of metrics, with some "strategic" metrics directly causal to O&S Cost, and other "subordinate" metrics that are causal to the strategic metrics. The leverages (λ_i's) to be estimated in Step 4 of the Metrics Thermostat are obtained from the slope coefficients

from multiple regression analyses of the relationships indicated in the diagram. Each arc in the diagram represents a slope coefficient to be estimated by regression analyses. The dependent variable in each regression is the construct with arrows leading into it, while the independent variables are the subordinate, causal metrics that have arrows pointing into the dependent variable. A metric's leverage with respect to O&S Cost is the sum of the products of all the slope coefficients for the arcs along all the paths leading from the metric to the supra-ordinate metric, O&S Cost. For example, in the diagram, the constructs System Manpower Requirements, System Training Requirements, Ease of System Upgrade/Technology Insertion, Software Support and Maintenance Requirements, and Corrective Maintenance all point directly to O&S Cost. Therefore, O&S Cost is regressed on these metrics and a slope coefficient (if it is statistically significant) is assigned to each arc connecting the metrics directly to O&S Cost. Next, the metrics that were the independent variables in this regression and have arrows leading into them are regressed on the appropriate subordinate metrics. For example, System Manpower Requirements is regressed on the metrics for Degree of Automation and Corrective Maintenance. The total leverage for Corrective Maintenance is its slope coefficient with respect to O&S Cost from the first regression (direct relationship to O&S Cost) plus the product of its slope coefficient with respect to System Manpower Requirements and the slope coefficient of System Manpower Requirements with respect to O&S Cost (indirect relationship to O&S cost). Moving left across the diagram and down the hierarchy of metrics, the leverage of Inherent Reliability with respect to O&S Cost, is its slope coefficient with respect to Corrective Maintenance, times the total leverage of Corrective Maintenance with respect to O&S Cost. These leverages are the results presented in the following chapter.

Figure 4-1 Causal Diagram: Hierarchy of Metrics
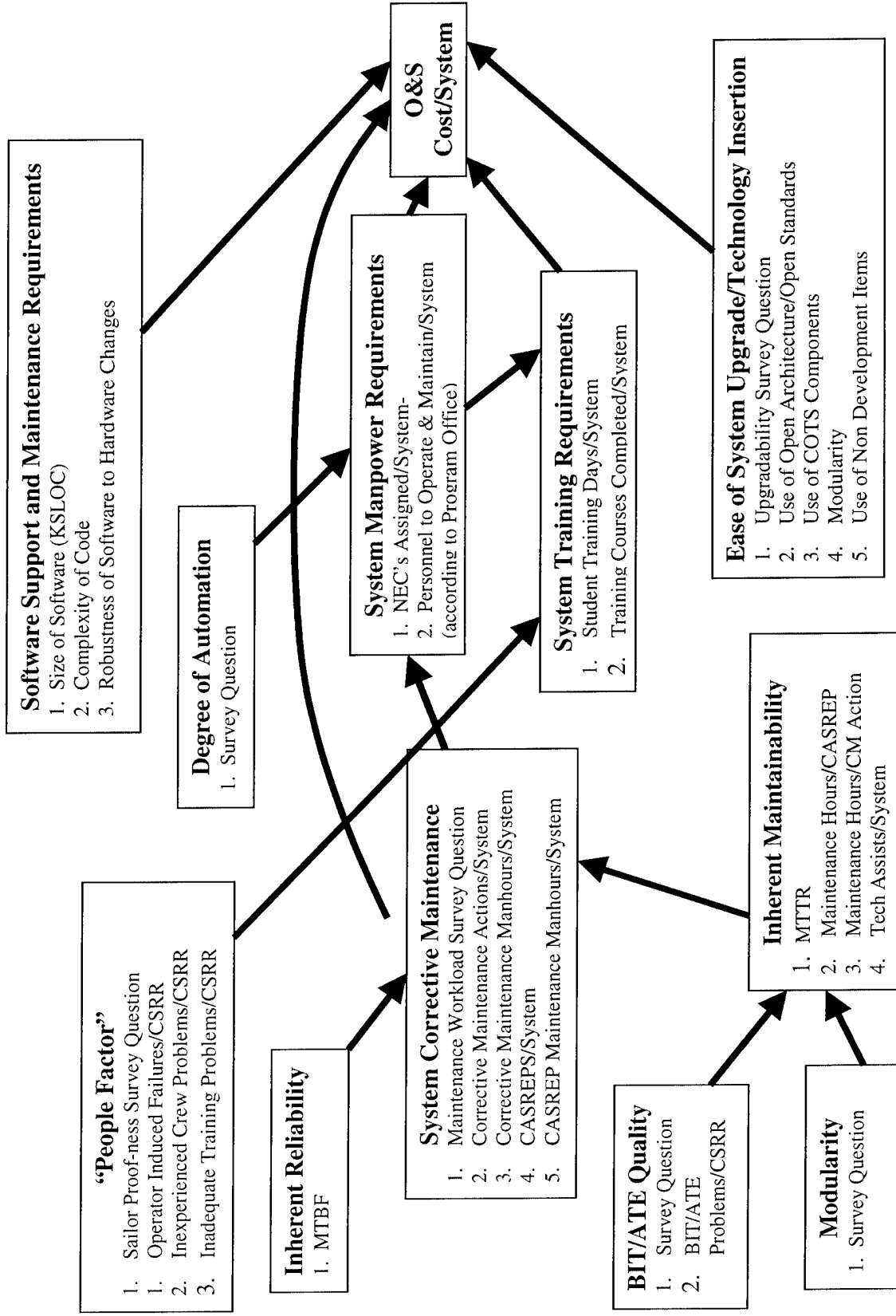
5-1   Overview

The 4th step in the Metrics Thermostat entails using multiple regression analysis to estimate the leverage (λ_i) for each metric, with respect to the appropriate supra-ordinate goal(s). This chapter describes the implementation of this step with the data and metrics described in the previous chapter.

Section 5-2 addresses the quality of the data set used in this analysis regarding the reliability of the data, the sparseness of the data set, and the volatility of the cost data. This section details the necessary *caveats* and remedies (when applicable) antecedent to a regression analysis of the data.

Section 5-3 discusses the general statistical methodology applied to the data set, beginning with the hypotheses depicted in the causal diagram in Figure 4-1.

Section 5-4 presents the results of each of the regressions applied to the data.

5-2   *Data Quality and Considerations*

This section addresses the three major issues affecting the quality of the data set (and thus, that of the data analysis); the reliability of the data, the sparseness of the data set due to data that was not available, and the volatility of the O&S cost data.

5-2-1 Reliability of the Data

Reliability refers to the internal consistency of data. A valid metric is one that is a good (i.e. precise and accurate) measure of the underlying construct it is supposed to measure. For metrics to be valid, they must be reliable. (Note that reliability does not imply validity, but rather, a lack of reliability implies a lack of validity. Metrics can be both reliable and wrong if they are consistent measures of the wrong thing.) Complementary metrics (as defined in Section 4-4-3) are metrics that are intended to measure the same underlying construct or concept. For example, the metric "Personnel Assigned NEC's per System" and the metric "Personnel per System According to the Program Office" are both intended to measure a system's manpower requirements. If these two metrics are shown to be unreliable (i.e. inconsistent with each other), then one (or both) of them is (are) not valid measure(s) of system manpower requirements. On the other hand, if these metrics are shown to be reliable, then their sum (or average) can be used as a reliable metric for manpower requirements. Summing (or averaging) two or more reliable complementary metrics provides an advantage in that the reliability of the sum (or average) of multiple reliable metrics can be greater than any one metric by itself. Thus, it is possible to create a new "purified" metric with greater over-all reliability by summing or averaging reliable complementary metrics. Therefore, the term purified metric, coined by LaFountain in his application of the Metrics Thermostat at Xerox Co., is used to refer to metrics created by summing or averaging reliable complementary metrics (LaFountain 1999).

A popular measure of reliability is Cronbach's alpha (Cronbach 1951). For a given set of metrics that are supposed to measure the same construct, Cronbach's alpha compares the degree to which the metrics co-vary to the total variance of the data set. The more the complementary metrics co-vary with each other as a fraction of the total variance, the greater the magnitude of Cronbach's alpha, and thus, the greater the internal consistency of the data. Cronbach's alpha, like the more commonly known Pearson's correlation coefficient, varies from –1 to +1. The greater the absolute value of Cronbach's alpha, the greater is the internal reliability of the data. While the literature does not provide an absolute minimum threshold for reliability, in the preliminary stages of research,"modest reliability in the range of 0.5 to 0.6 will suffice" (Peter 1979). In most cases, the data used in this research met or exceeded this threshold.

In addition to assessing the reliability of complementary metrics in creating purified metrics, Cronbach's alpha was used to assess the internal consistency of the metrics that were collected by administering a questionnaire to those who manage and maintain the systems in the study (recall the brief discussion about this in Section 4-2, pages 40-41). In total, 12 of the metrics in the data set were in the form of survey questions. To filter out some of the subjectivity of these measures, the survey questions were administered (whenever possible and appropriate) to three different individuals from the organizations that support and manage the systems in the data set;

one from the program office, one from the ISEA, and one "waterfront" technician from FTSCLANT. Of the 12 survey questions, 7 pertain to the technical design attributes of the system and these were asked only of the program manager and ISEA. The remaining 5 questions pertained to the system's performance and maintenance attributes and these were asked of the FTSCLANT technicians in addition to the ISEA and the program office. (Note, that the respondent from the program office, if not an engineer himself or herself, would typically defer the more technical questions to someone in the program office of an engineering background). To assess the consistency of the survey questions, Cronbach's alpha was calculated for the different responses for each question across the data set. The following tables present the reliability of each question.

Table 5-1 – Reliability of Design Attributes Survey Questions*

Metric	Cronbach's alpha
Use of Non-Development Items	0.10
Use of COTS Components	0.41
Use of Open Architecture-Open Standards	0.55
Upgradability/Ease of Technology Insertion	0.42
Thousands of Lines of Source Code (KSLOC)	0.95
Software Complexity	0.79
Robustness of Software to Hardware Changes	0.50

*These questions were answered (whenever possible) by a representative of the program office and the ISEA.

Table 5-2 – Reliability of Operation & Maintenance Survey Questions**

Metric	Cronbach's alpha
Modularity of the System	0.59
Degree of Automation	0.72
System Corrective Maintenance	0.81
Quality of Built-In Testing	0.53
Sailor Proofness of the System	0.54

**These questions were answered (whenever possible) by a representative of the program office, the ISEA, and a technician from FTSCLANT.

For the most part, the survey-metrics exhibited acceptable reliability for preliminary research. The low Cronbach's alpha for the question intending to measure the degree to which a system's design exploits Non-Development Items (NDI) revealed that the question was lacking in internal consistency. Therefore, this metric was excluded from the analysis. Two other metrics, "Use of COTS Components" and "Upgradability/Ease of Technology Insertion" did not meet the minimum threshold of 0.5. However, these two metrics were retained for analysis because when summed with other complementary metrics, they contributed positively to the over-all reliability of the summed metric for "System Upgradability." Excluding the metric for use of NDI, the average reliability for the individual metrics was 0.58, and the median reliability was 0.55. When standardized and summed with complementary metrics, the reliability of the resulting "purified" metrics was even greater, typically around 0.7 (discussed in Section 5-4).

5-2-2 Sparseness of the Data Set

A second issue affecting the quality of the data set was the unavailability of data for some systems, for some metrics, during certain years. Data collection for this study began with the systems in Table 4.1 and compiled yearly data for the metrics described in Chapter 4, as far back as the VAMOSC reports provided cost data on them. Thus, an individual data point for this study consisted of data for a given system for a given year. Therefore, a data point refers to a system-year of data.

The main difficulty in obtaining complete system-years of data arose from the fact that the data were compiled from so many different sources. The MRDB had reliability (MTBF, not the reliability discussed in the previous section) and maintainability (MTTR) data on only 16 of the VAMOSC systems (though fortunately, the MRDB data that was available goes as far back, time-wise as the VAMOSC data in most cases). Data from the Navy TSP was available for all but 10 of the systems in the VAMOSC reports, but not before FY 1992 (the VAMOSC data goes as far back as FY 1986 for 12 of the systems in Table 4-1, however, approximately 73% (or 278 of 379 data-years) of the VAMOSC data comes from years after FY 1991). The following table shows the number of systems for which data was available from VAMOSC and the fraction of those systems for which Navy TSP, and the MRDB also supplied data for FY 1986 to FY 1999. (The fraction "7/12" in the third row of the right most column indicates that MRDB furnished data on 7 of the 12 systems for which VAMOSC maintains FY 86 data.)

Table 5-3 – Data Availability by Year and Source

	VAMOSC Data	TSP Data			MRDB Data
Year	Systems in the VAMOSC SSR	CASREP Data	TAVR	CSRR Root Causes	MTBF and MTTR
86	12	0	0	0	7/12
87	17	0	0	0	8/17
88	17	0	0	0	8/17
89	17	0	0	0	8/17
90	17	0	0	0	9/17
91	21	0	0	0	12/21
92	21	18/21	0	0	12/21
93	31	28/31	0	24/31	14/31
94	31	26/31	10/31	24/31	14/31
95	31	26/31	26/31	26/31	15/31
96	33	28/33	24/33	27/33	16/33
97	45	37/45	36/45	36/45	16/45
98	50	40/50	36/50	36/50	16/50
99	36*	26/36	26/36	25/36	10/26

* For FY '99 VAMOSC data for 14 systems was excluded from the analysis because the SSR were known to be missing some data elements.

In addition to data that was unavailable from the MRDB and the Navy TSP data sources, some of the program managers, ISEA's and FTSCLANT technicians were not available to provide inputs

for the 12 survey-based metrics. In a few cases, the systems in the VAMOSC reports were no longer in the fleet, and therefore, no longer had ISEA's or program offices. In other cases, the program offices or ISEA's could not be reached for input. The following table summarizes the availability of data from the program managers, ISEA's, and FTSCLANT technicians by system-year.

Table 5-4 – Survey Data Availability by Year and Source

	VAMOSC Data	Survey Questions				
Year	VAMOSC SSR's	FTSCLANT	ISEA	PM	PM or ISEA	PM AND ISEA
86	12	12/12	7/12	8/12	9/12	6/12
87	17	17/17	11/17	12/17	14/17	9/17
88	17	17/17	11/17	12/17	14/17	9/17
89	17	17/17	11/17	12/17	14/17	9/17
90	17	17/17	11/17	12/17	14/17	9/17
91	21	21/21	13/21	16/21	18/21	11/21
92	21	21/21	13/21	16/21	18/21	11/21
93	31	26/31	21/31	20/31	27/31	12/31
94	31	26/31	21/31	20/31	27/31	14/31
95	31	26/31	21/31	20/31	27/31	13/31
96	33	28/33	23/33	21/33	29/33	14/33
97	45	36/45	32/45	29/45	40/45	21/45
98	50	38/50	37/50	35/50	45/50	26/50
99	36	18/36	13/26	12/26	21/26	14/26

Given the amount of data that was missing, it was important to take precautions in the regression analysis. Six systems for which neither PM, nor ISEA inputs were available were excluded from the data analysis, reducing the total number of data points to 316. For the 11 survey questions that were used in the analysis, the average of the responses from the program office, the ISEA, and the FTSCLANT (when applicable) was used as the metric score for each metric. In cases where either the program office or the ISEA was not available, the average without the missing input was used (this was the case in 134 of the 316 remaining data points used in the analysis).

A similar approach was used whenever one of two or more complementary metrics with sufficient reliability (0.5 or higher) had missing values.

Otherwise, missing values were excluded pair-wise or list-wise or were replaced with the mean value for the respective metric, depending on how much data was missing. Since different regressions used different subsets of the data, the details of how missing values were handled in each case are deferred to Section 5-3.

5-2-3 Volatility of the Operating and Support Cost Data

A third factor about the data that was important to consider before launching into regression analysis was the time-volatility of the cost data. While O&S cost varies among different types of systems, it may also vary greatly within the different phases of life for the same system. For example, when a system is first introduced to the fleet, it may require costly engineering changes

and modifications as unforeseen problems arise in its initial operation. Furthermore, in the early, procurement stages of the system's life, the O&S cost on a per system basis may be inflated if the number of systems in the fleet is small enough (even though the Navy may be spending large amounts of money on the program as a whole). As the system matures, most of the early problems will have been resolved and it usually costs less to operate and maintain in steady state; until such time as wear and tear on the system necessitate major corrective maintenance or overhaul. Finally, in the latter years of a system's life, O&S costs tend to taper off as funding for things such as maintenance, training, and upgrades diminishes. Moreover, the Navy will upgrade or overhaul its systems every so often causing periodic spikes in cost that must also be accounted for.

As expected, the systems in the VAMOSC database exhibited quite a bit of volatility over time. For example, one sonar system incurred O&S costs that differed by a factor of 9.7 in different years of its life. Even in two consecutive years of this same system's life, O&S costs sometimes differed by as much as a factor of 2. This kind of variability in cost could give a misleading representation of a system's true O&S cost if only a few years of data were available (as was the case for many of the systems in the study). The potential for an inaccurate estimate of a system's cost would be even greater if the system were in only one phase of its life during the two or three years for which data is available. For a concrete example, consider the following chart of population and O&S cost trends for one particular system.

Figure 5-1 – Volatility of O&S Costs
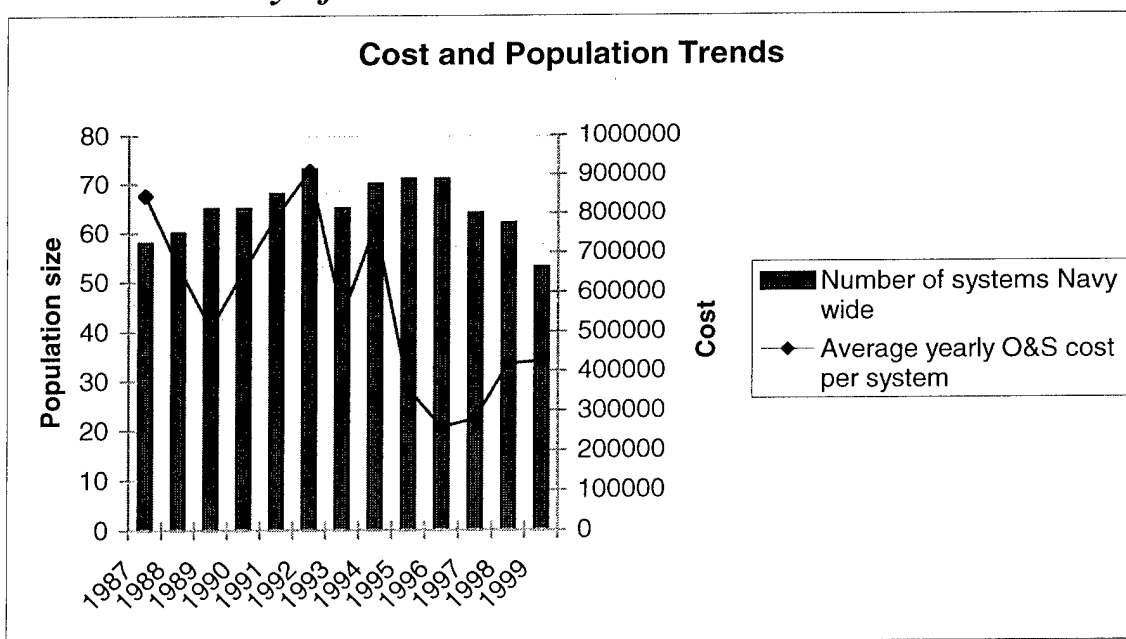
For this system, the population and cost trends suggest that the system was in two distinct phases of its life for the years shown. For the years 1987 to 1994, the system's O&S cost wavered around $700K per system while the population size increased slightly. Around 1996, the population started to decline as the system entered the phase out period of its life. This downward trend in population size coincided with a steep drop-off in cost (approximately half of

the cost in the previous phase). When compared to a similar system in a different phase of life, one might incorrectly conclude that this system is less (or more) costly than the other system when the difference is actually attributable to life cycle phase. Furthermore, there are many systems in the study for which only the last few years of data are available. Consider what a misleading representation of the O&S cost of the system in Figure 5-1 one would have if only the last 5 years of data ('95 – '99) were available.

In a conversation with an analyst from VAMOSC, the analyst indicated that most people who use VAMOSC data for analysis consider about a seven- year window of data to be sufficiently broad to estimate a system's O&S cost. One approach to accounting for volatility would be to include in the study only those systems with at least 7 years of data, taking the average O&S cost per system over the years of available data. Unfortunately, this would mean excluding 26 of the systems in the data- base. Additionally, merely taking the average would not allow the possibility of controlling for what phase of life cycle the systems were in during the years for which data is available. Therefore, I decided to categorize each year of data for each system according to what phase of life the system was in during that year and use this as a "dummy variable" in my statistical analysis.

The DoD's Cost Analysis Improvement Group (CAIG) addresses the issue of O&S cost varying over a system's life in its "Operating and Support Cost Estimating Guide." According to CAIG, there are three distinct phases to a system's life cycle to consider when estimating O&S costs: phase in, steady state, and the phase out.

Figure 5-2 – Life Cycle Phases (CAIG 1992)
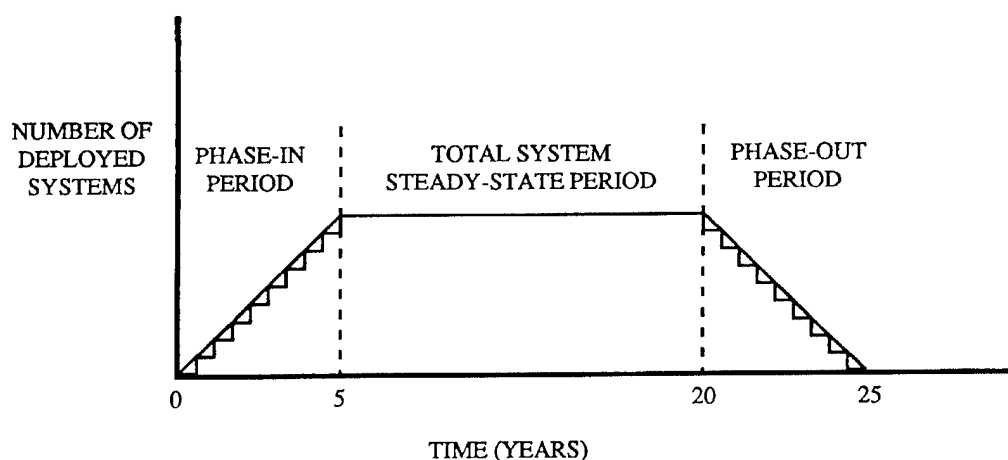
In addition to the three life cycle phases mentioned above, one must also account for the fact that the Navy will periodically modernize or overhaul a system. These activities include such things as major upgrades, overhauls, or the purchase of replenishment spares or new equipment. Since the expenses for modernization and/or overhaul are often concentrated in one or two fiscal years, these activities will often cause a spike in a system's O&S cost data, as seen below in Figure 5-3.

65

Figure 5-3 – O&S Cost Spike
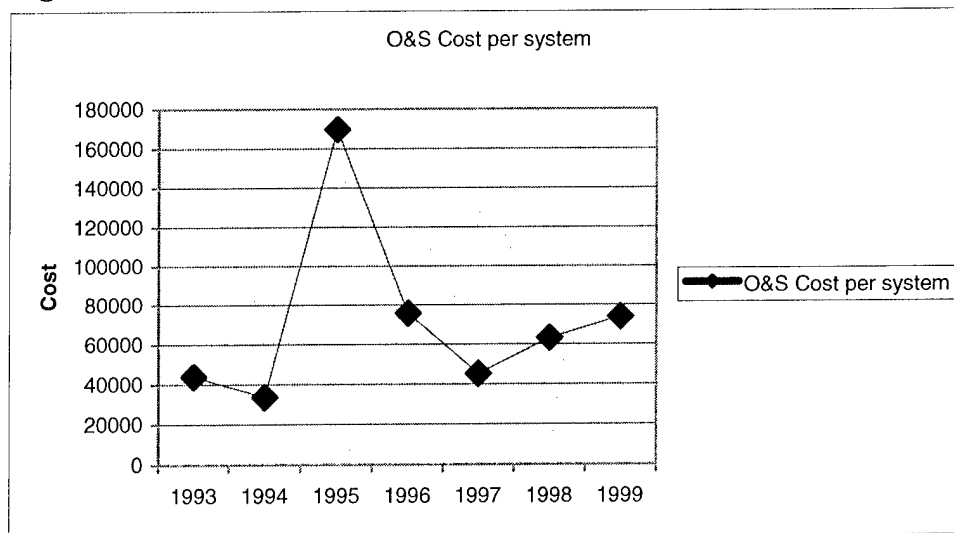
For the years 1993 to 1999, this system's O&S cost per system remained fairly constant, except for the year 1995, for which the O&S cost almost tripled the "normal" O&S cost. As it turns out, VAMOSC reports a substantial expenditure for "System Component Rework" during this year that itself is more than double the total O&S cost for the system in any other year from 1993 to 1999. If not accounted for in the statistical analysis of the data, such a cost spike might produce misleading results. For example, if one were to calculate a Pearson's correlation coefficient for the above system to measure the correlation of the number of times the system broke down per year with its O&S cost per year, without taking the spike into account, the coefficient would be –0.126. This would suggest that as the number of times the system breaks down increases, the yearly O&S cost decreases. However, if one takes the cost spike into account, the correlation coefficient becomes +0.729, clearly a more reasonable result, as one would expect a positive correlation between cost and system failures. Since almost every system in the study had similar cost spikes, I added a fourth category to the three described by CAIG to account for years for which a system has a cost spike attributable to modernization efforts, overhauls, or other "one time" activities.

To make the process of categorizing each year of data for each system more objective, I established the following criteria before categorizing each year of data for each system:

66

- Phase-In Period:
  - o Significant boost/upward trend in overall population size, system still in production during these years
- Steady State Period:
  - o No major change in population size and
  - o No major upgrade or overhaul costs incurred
- Phase-Out Period:
  - o Significant decline in overall population size
- Major Upgrade/Overhaul Year:
  - o Any year not fitting the above criteria during which there were major upgrades, overhauls, ORDALT's or other "curve balls" that would cause a salient spike in the system's cost data.

The first step I took was to make charts of each system's population and cost trends like the one in Figure 5-1. From the graphs, I noted the population and cost trends. This helped me to form an initial idea of each system's status during the years for which data was available. For some systems, the categorization of system life cycle phases were immediately clear. For other systems, like the one in Figure 5-4, it was much less obvious what phase of life they were in.

Figure 5-4 – Surface Sonar Life Cycle Cost Spikes
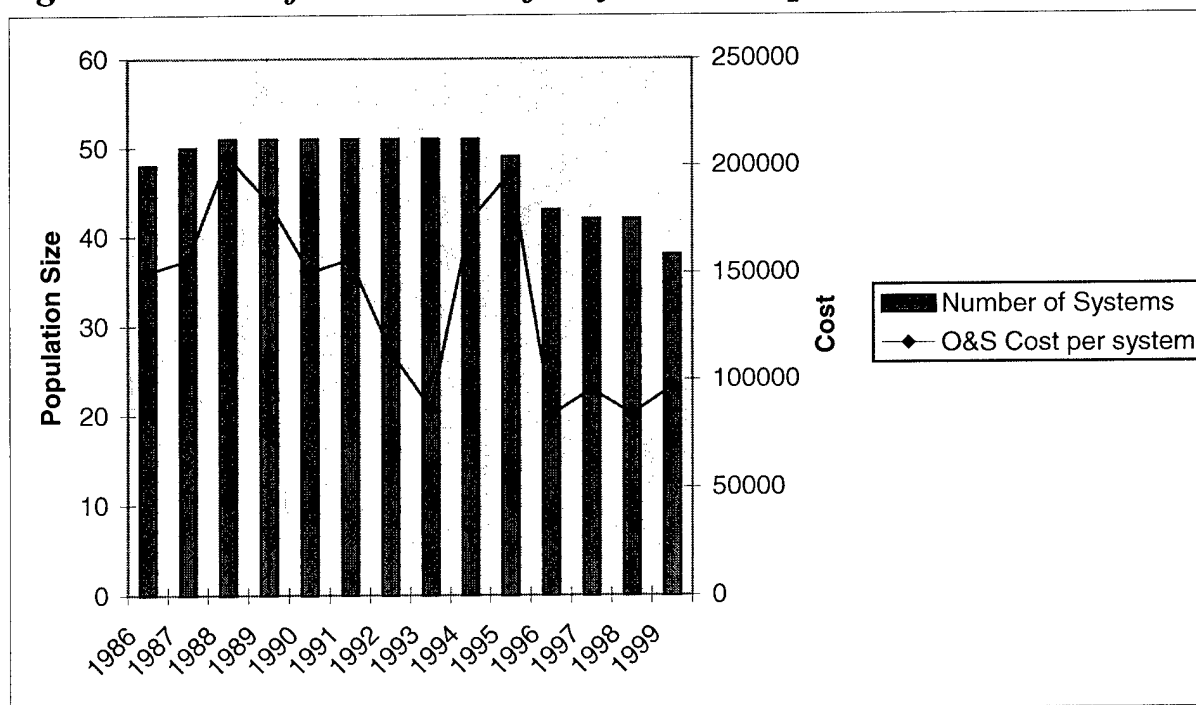
After combing through the data for all the systems, I noticed that whenever a system had a particularly salient cost spike, I would almost always find a significant expenditure in the VAMOSC data in one or more of the following data fields:

1002.3.2 Fleet Modernization
1002.3.3 System Component Rework
1002.3   Engineering Technical Services
1002.4   Embedded Computer/Software Support

Generally, if costs from these categories added up to 40% or more of the O&S cost for a system for a particular year, there would be a spike such as the two in Figure 5-5 for the years 1988-1989 and 1994-1995.

After inspecting a system's cost and population data and trends, I would make an initial categorization of each year of available data according to the criteria mentioned above. For the systems that were difficult to categorize, the next step was to present the data, along with my initial attempt to classify life cycle phases, to the system's Program Manager or ISEA. The PM or ISEA would provide feedback and historical information about the systems that I would use to validate or improve my categorization. Often the PM or ISEA would give an explanation for cost spikes in certain years. For instance, the Program Manager for the sonar system in Figure 5-4 explained that a new COTS subsystem was installed around 1988 and upgraded around 1994.

In addition to making sure my categorizations were accurate, presenting the data to the PM's and ISEA's afforded me the opportunity to validate the data supplied by VAMOSC. In most cases, the PM's and ISEA's familiar with the systems were able to verify from their own recollection, the population and cost trends in the data from VAMOSC.

5-2-4 Final Data Set

After excluding those systems for which insufficient data was available, the following systems remained for analysis:

Table 5-5 – Listing of Systems Included in Regression Analysis

FY86-94	5"/54 CALIBER MK-42 GUN
FY86-99	5"/54 CALIBER MK-45 GUN
FY93-99	AN/BPS-15 SERIES(A-D) RADAR
FY97-99	AN/BPS-16 (V) RADAR
FY86-99	AN/BQQ-5 SONAR SYSTEM
FY97-99	AN/BQQ-6 SONAR
FY97-99	AN/BQS-15 SONAR DETECTING-RANGING SET
FY93-99	AN/BRD-7 AND 7A ELECTRONIC COUNTERMEASURE SET
FY98-99	AN/SLQ-48(V) NEUTRALIZATION SYSTEM MINE
FY98-99	AN/SPS-40B RADAR
FY98-99	AN/SPS-40E RADAR
FY93-97	AN/SPS-40C/D/E
FY93-99	AN/SPS-48C RADAR
FY93-99	AN/SPS-48E RADAR
FY87-99	AN/SPS-49 RADAR
FY86-99	AN/SPS-55 RADAR

FY91-97	AN/SPS-64 (V) 3 AND 9 RADAR
FY98-99	AN/SPS-67 (V) 1 RADAR
FY98-99	AN/SPS-67 (V) 3 RADAR
FY91-97	AN/SPS-67 (V) 1 & 3 RADAR
FY91-99	AN/SQQ-89 SURFACE ASW COMBAT SYSTEM
FY85-99	AN/SQS 53A SONAR
FY86-99	AN/SQS 56 SONAR
FY93-99	AN/SYQ-20 ADVANCED COMBAT DIRECTION SYSTEM
FY93-99	AN/SYS-2 INTEGRATED AUTOMATIC DETECTION AND TRACKING SYSTEM
FY96-99	AN/USC-38 EHF SATCOM
FY97-99	AN/WLQ-4 (V)/(V) 1 COUNTERMEASURE RECEIVING SET
F797-99	AN/WLR-8 (V) 2/ (V) 5 COUNTERMEASURE RECEIVING SET
FY86-99	CLOSE-IN WEAPON SYSTEM MK-15
FY86-99	COMBAT CONTROL SYSTEM MK-1
FY97-99	COMBAT CONTROL SYSTEM MK-2
FY87-99	HARPOON WEAPON SYSTEM
FY93-99	MK 14 WEAPONS DIRECTION SYSTEM
FY97-99	MK 23 TARGET ACQUISITION SYSTEM (TAS)
FY87-99	MK 26 GUIDED MISSILE LAUNCHING SYSTEM
FY87-99	MK 41 VERTICAL LAUNCHING SYSTEM
FY97-99	MK 57 NATO SEA SPARROW SURFACE MISSILE SYSTEM
FY93-99	MK 74 MISSILE FIRE CONTROL SYSTEM
FY97-99	MK-75 76MM GUN OTO-MELARA
FY87-99	MK-86 GUN FIRE CONTROL SYSTEM
FY91-99	MK-92 FIRE CONTROL SYSTEM
FY96-99	MK-116 UNDER WATER FIRE CONTROL SYSTEM MOD 1/2 AND 4
FY86-99	MK-117 FIRE CONTROL SYSTEM
FY97-99	MK 118 UNDERWATER FIRE CONTROL SYSTEM (UFCS)
FY97-99	RAM MK 31 GUIDED MISSILE WEAPONS SYSTEM

This section describes the general procedures and techniques used in the analysis of the data. The results of the analysis are presented in Section 5-4.

Analysis began at the top of the metrics hierarchy depicted in Figure 4-1. Starting with O&S Cost per System, each metric was regressed on those lower-level metrics hypothesized to have a causal relationship with the metric. Initially, there were 5 regressions to perform. (This number would later be reduced to 3 for reasons described in the next section.)

Regression 1: Operating and Support Cost per System
Regression Variables:
  System Manpower Requirements
  System Training Requirements
  Ease of System Upgrade/Technology Insertion
  Software Support and Maintenance Requirements
  Corrective Maintenance

Regression 2: System Training Requirements
Regression Variables:
  People Factor
  System Manpower Requirements

Regression 3: Manpower Requirements
Regression Variables:
  Degree of Automation
  Corrective Maintenance

Regression 4: Corrective Maintenance
Regression Variables:
  People Factor
  Inherent Reliability
  Inherent Maintainability

Regression 5: Inherent Maintainability
Regression Variables:
  Modularity
  BIT/ATE Quality

Before performing each regression, it was necessary to construct the regression variables (or "purified metrics") from their constituent complementary metrics. Whenever two or more complementary metrics were used to measure the same construct, reliability analysis was performed on the constituent complementary metrics in order to verify that they were internally consistent (i.e. measured the same thing). This began by looking at a correlation matrix with the complementary metrics and the dependent variable for that particular regression. The correlation matrix revealed the strengths and signs of the inter-correlations among the complementary

70

variables in addition to their correlations with the dependent regression variable. If one of the complementary metrics had a low correlation with the other metrics, then it was identified for potential elimination from the analysis. Those complementary metrics with the strongest inter-correlations (internal consistency) and correlations with the regression dependent variable were the best candidates for incorporating into the purified metric.

Next, Cronbach's alpha was computed for the complementary metrics from which the regression variables would be created. (Cronbach's alpha actually measures the reliability of the purified metric that is created by summing all the complementary metrics. Moreover, Cronbach's alpha assumes that the metrics all have the same scale of measurement, so the metrics had to be standardized before regression analysis.) The statistical software package used provides a Cronbach's alpha score for the reliability of all of the metrics together. In addition, for each complementary metric, the package calculates an alpha score indicating what the reliability of the sum of all the other metrics would be if that metric were excluded from the summed purified metric. If the overall reliability of the purified metric could be significantly improved by excluding a constituent metric, then the constituent metric was excluded from the purified metric.

The purified metrics were then constructed by averaging those constituent complementary metrics that were retained after reliability analysis (after they had been standardized, since their measurement scales were not always the same). (Complementary metrics were averaged rather than summed because the software package used in the analysis automatically skips missing values when computing the mean of a set of variables, whereas computing the sum of complementary metrics would have required adjusting the sum on a case-wise basis for missing values. Statistically, the mean and the sum differ by only a constant, and are equivalent for use in regression analysis. However, using the average rather than the sum saved the trouble of adjusting the sum for missing values on a case-wise basis.)

Once the purified metrics were constructed, they could be used as variables in the regression analysis. Data points for FY 1998 were withheld from the analysis in order to evaluate the predictive abilities of the models, in addition to verifying that the models were not over-fitting the data set. Data points with missing values for the purified metrics were handled in the following way. If a metric was of high reliability and did not have missing values for a large fraction of the data points, then the missing values for that metric were filled with the grand mean for that metric (across all available data points for that metric). If, on the other hand, there was a large fraction of data points for which a purified metric was lacking data, then missing values were either excluded list-wise or pair-wise depending on the particular regression. In most cases, all three alternatives were explored and the slope coefficients, significances, and predictive abilities of the alternatives were compared.

With each regression, the standard assumptions regarding residuals were verified using Normal P-P plots, histograms, and scatter plots of the residuals with the dependent and independent variables. (Most of the details about the residuals are left to the Appendix. Only the most important details of the residual analysis are mentioned in this chapter.) Studentized residuals, leverage, and standardized DFFITS statistics were examined to identify influential data points. Data points with studentized residual greater than 3 in absolute value, leverage greater than $3p/n$ (where p is the number of slope coefficients estimated and n is the number of data points), or

very large Standardized DFFITS were flagged for investigation (Welsh 1999). Once influential data points were identified for investigation, they were only excluded if a compelling reason to eliminate them could be found in the data. Typically, highly influential data points would corresponded to a spike in the cost data or the very beginning or end of a system's life cycle in which the system's small population size inflated the systems O&S Cost per system. In each case where influential observations were excluded from the analysis, the model's adjusted R^2 improved as well as the model's predictive ability on the withheld data.

This section documents the specific regression analyses and their results, beginning at the top of the metrics hierarchy in Figure 4-1 and progressing down the hierarchy to the left-most metrics in the causal diagram.

5-4-1 Regression of Operating and Support Cost per System

Five metrics were initially hypothesized to drive system O&S cost directly:

- System Manpower Requirements
- System Training Requirements
- Corrective Maintenance
- System Software Support and Maintenance Requirements
- The Ease of System Upgrade/Technology Insertion

System Manpower Requirements:
General Remarks: The VAMOSC data only tracks the costs of enlisted personnel assigned to systems. It does not include the manpower costs incurred by the officers that must supervise them or the enlisted personnel who are not technically assigned to the ship. Thus, it is likely that the true cost of a system's manpower requirements is higher than the cost estimated by this analysis.

Reliability: The two constituent metrics used to measure System Manpower Requirements were Personnel Assigned NEC's per System and Personnel per System According to the Program Office. The reliability of these two metrics together was a modest 0.51. This modest reliability is most likely attributable to differences described in Section 4-4-2. Both metrics have respective strengths and weaknesses and both were used to measure manpower as the two metrics complement each other well and have sufficient reliability for preliminary analysis.

System Training Requirements:
General Remarks: For some systems training can require several weeks, even months. Before recent policy changes to reduce the amount of training given to sailors, training alone could account for a large fraction of a sailor's first enlistment. For every sailor the Navy trains, the Navy must pay for the salaries and benefits of the trainees, as well as those of the personnel who operate the training facilities. Training, therefore, has a direct contribution to system O&S Cost. On the other hand, one might also argue that quality training may actually decreases system O&S cost by reducing the amount of system operator or maintenance induced failures and increasing the quality of preventative maintenance. However, the data did not reveal this.

Reliability: The two constituent metrics for System Training Requirements, Student Training Days per System and Training Courses Completed per System were of such poor reliability (0.24) that they could not be used together to measure System Training requirements. This is most likely attributable to the metric Training Courses Completed. This measures only the number of training courses completed for a system, without measuring the length of the training or the number of personnel trained. Student Training Days per System is likely a more valid measure of System Training Requirements. Moreover, Student Training Days per System correlates

strongly to the other manpower related metrics. When included with the manpower metrics Personnel Assigned NEC's per System and Personnel per System According to the Program Office, the over-all reliability of the metrics increases to 0.65. This suggests that Student Training Days should be grouped with these metrics as measures of System Manpower and Training Requirements. Thus, System Manpower Requirements and System Training Requirements were consolidated into one more reliable metric, System Manpower and Training Requirements.

System Corrective Maintenance:
General Remarks: The corrective maintenance generated by a system has a direct effect on O&S cost in the form of spare parts and consumables necessary to maintain the system, as well as an indirect effect on O&S cost in that maintenance also drives System Manpower Requirements. Thus, the total effect of a system's maintenance is the direct effect, plus the indirect effect it has on cost via manpower.

Reliability: The constituent complementary metrics of System Corrective Maintenance were as follows:

- Maintenance Workload Survey Question (asked of program office, ISEA, and FTSCLANT technicians)
- Number of Corrective Maintenance Actions per System per Year
- Corrective Maintenance Man-Hours per System per Year
- Number of CASREPs per System per Year
- Number of CASREP Maintenance Man-Hours per System per Year

A correlation matrix (see Appendix 2) and reliability analysis for these metrics suggested that the Number of CASREPs per System was not consistent with the other Corrective Maintenance metrics (moreover, it did not correlate strongly with cost). This is easy to explain given the way CASREPs are reported. Systems that are more critical to the mission of a ship (but not necessarily maintenance intensive) are more likely to be "CASREP'ed" when they fail than others (therefore, the number of CASREPs/System may depend more on the judgement of the crew and the mission essentialness of the equipment, rather than the maintenance generated by CASREP failures). Without this metric, the Corrective Maintenance metrics have an over-all reliability of 0.71.

Software Support and Maintenance Requirements:
Reliability: The constituent complementary metrics for Software Support and Maintenance Requirements were:

- Thousands of Lines of Source Code (KSLOC)
- Survey Question on Complexity of Code (asked of program office and ISEA)
- Survey Question on the Robustness of the Code to Changes in Hardware (asked of program office and ISEA)

Together, these metrics had a reliability of 0.59.

<u>Ease of System Upgrade/Technology Insertion:</u>
Reliability: As previously mentioned, Use of Non-Development Items did not have adequate reliability for inclusion in the metrics designed to measure the degree to which a system lends itself to easy, affordable upgrades. All of these metrics were in the form of survey questions administered to system program managers and ISEA's. Without Use of NDI, the following metrics had a reliability of 0.73:

- Upgradability Survey Question
- Use of Open Architecture/Open Standards
- Use of COTS Components
- Modularity of System

<u>Binary Variables:</u>
Life Cycle Variables:
Binary "dummy" variables were used to represent the life cycle phase of the system for each system-year of data. As discussed in Section 5-2-3, each year of data for each system was categorized as Phase-In, Steady State, Phase-Out, or Major Upgrade/Overhaul. These dummy variables were intended to help account for some of the volatility of the cost data (recall the discussion in Section 5-2-3). The variables Phase 1, Phase 2, and Phase 3 correspond to the Phase-In, Steady State, and Phase-Out life cycle categories respectively. Whenever all three of these are equal to zero, then the data-year was in the Major Upgrade/Overhaul category.

Regime Variables:
Many of those interviewed along the data trail suggested that DoD and Navy policy changes in the early 90's had emphasized reducing DoD spending and that these policy changes had reduced both costs and readiness. Fiscal Year 1994 was the first full year of a new regime in the upper echelons of the DoD and also the year of congressional elections that may have changed to political climate with respect to government spending. The Perry Memo was published at this time, and a cursory examination of the VAMOSC reports suggests that less money was appropriated for upgrades and modifications in the later years of the reports. Quality of life initiatives to reduce the amount of preventative maintenance performed on ship and the amount training given to sailors also began around this time period.

The following box plots of O&S cost by year and regime (pre and post FY '94) corroborate what many suggested in interviews at NAVSEA and FTSCLANT.

Figure 5-5 – Box Plots of O&S Cost by FY
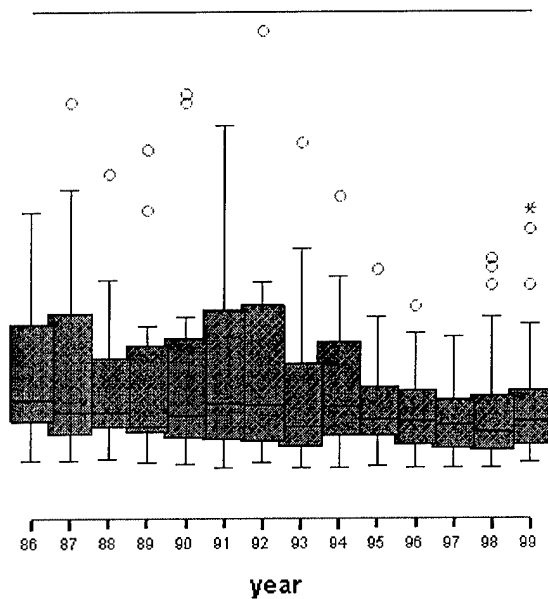
Figure 5-6 – Box Plots of O&S Cost by Regime
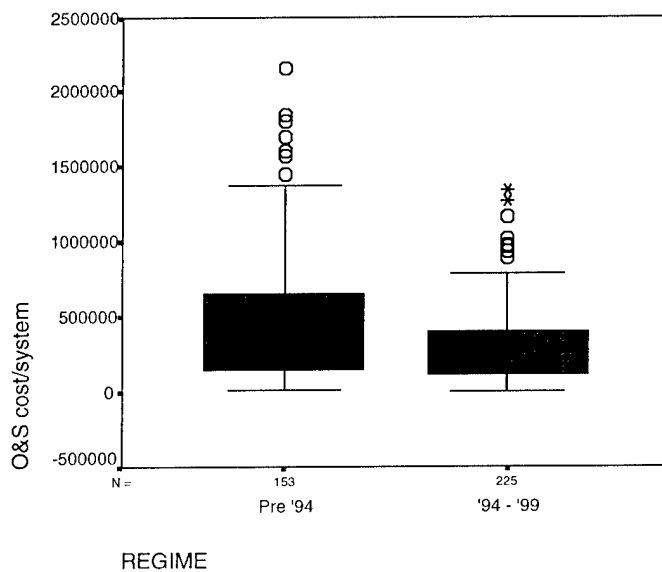
The box plots show that on the whole, O&S costs decreased after FY '94. The greatest differences apparent in the plots are in the upper percentiles of the box plots. This is most likely due to the fact that less money was spent upgrading, modifying, and acquiring new systems (as these expenditures typically correspond to the cost spikes previously discussed). The median and lower percentiles for the two regimes are closer, though they are also lower in data points from FY '94 and beyond. This is likely a reflection of the fact that O&S cost is as much a policy

76

decision as it is an outcome of the variables in this regression. On the other hand, this could also result from the fact that the systems for which VAMOSC keeps data have changed over time so that the systems in the data points before FY '94 are not all the same as those in FY '94 and afterwards. The following frequency diagram indicates that before FY '94, there was a relatively greater percentage of systems in the data set that were in the more expensive phases of their life-cycle (Build Up and Major Overhaul/Upgrade) than in years after FY '94.

Figure 5-7 – Life Cycle Phases by Regime
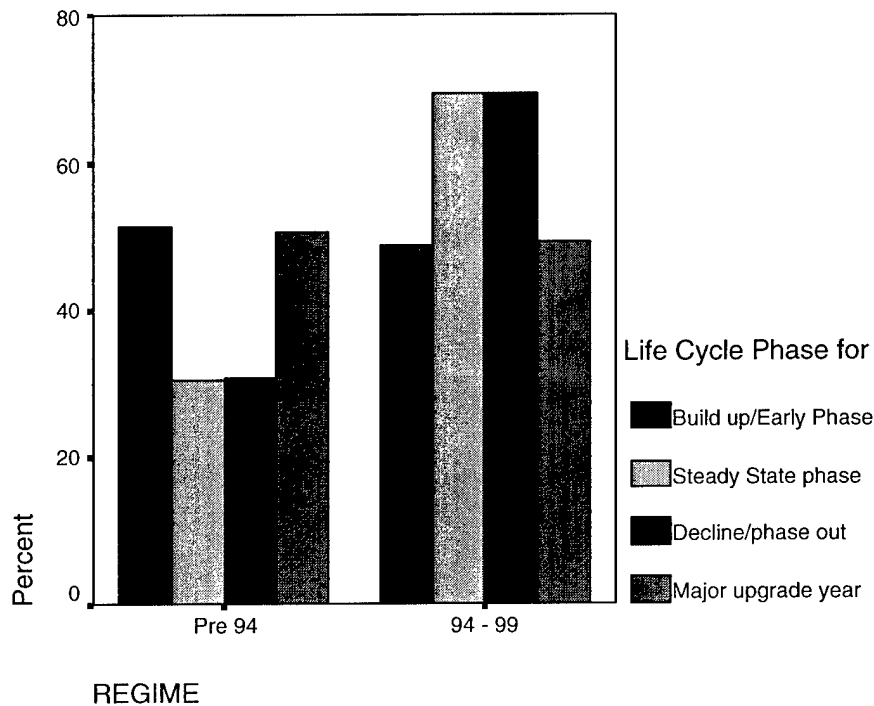
Whether because of policy changes or because additional systems were added to the VAMOSC database in the 90's, an ANOVA Table also confirms that there indeed was a significant difference in O&S costs between the two regimes.

Table 5-6 – ANOVA Table for Regime (Pre 94 and 94-99)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.15387E+12	1	1.15387E+12	7.813013331	0.005677356
Within Groups	3.04233E+13	206	1.47686E+11		
Total	3.15772E+13	207			

Therefore, a binary variable indicating whether the data point was from before FY '4 or FY '94 and beyond was added to the model, in addition to the life cycle variables to account for some of the volatility of the cost data.

As an alternative to this somewhat subjective process of categorizing the data points according to the system's life cycle phase in each year of data, a similar regression model was attempted without the life cycle variables. In order to "smooth out" the volatility of the cost data, the data for each system was averaged and used as the data points for that system in an implicit weighted least squares regression. The number of data points for each system was retained in this model, with each data point for a given system containing the mean data for that system rather than the yearly data. Thus, weights were implicitly assigned to data points according to how many years of data were available for the system in question. A system with 4 years of data still had 4 identical data points, and a system with 7 years of data had 7 identical data points, and so on.

Therefore, there were two alternative models for comparison.

1. Model 1: With yearly data, binary life cycle variables and a binary variable for "Regime".
2. Model 2: Averaged Data – each system's data averaged and used for as many data points as there were years of data for the system in the original data.

The following table compares the two alternative models in terms of their explanatory power with respect to the data set, their predictive abilities with respect to the withheld data from FY 1998, and the slope coefficients of each metric.

Table 5-7 – Regression Results for O&S Cost, Models 1 and 2

	Model 1	Model 2
Explanatory Power (Data Set)		
R^2	0.682	0.697
Adjusted R^2	0.673	0.693
Significance of Model	0.000	0.000
Predictive Ability (FY 98 Data)		
PRESS	6.31E+11	9.24E+11
Squared Correlation (R^2)	0.728	0.698
Standardized Coefficients with Significance		
Manpower & Training	+0.564 (.000)	+0.597 (.000)
Corrective Maintenance	+0.233 (.000)	+0.201 (.000)
Ease of Upgrade/Technology Insertion	-0.055 (.142)	-0.045 (.219)
Software Support and Maintenance	+0.068 (.128)	+0.166 (.000)
Phase 1	-0.200 (.000)	N/A
Phase 2	-0.293 (.000)	N/A
Phase 3	-0.399 (.000)	N/A
Regime	-0.199 (.001)	N/A

The results of the two models are fairly consistent. There is little difference in the explanatory power of the two models. In with both models, the squared correlation of the predicted values with the withheld values is actually higher than the R^2 for the data used in the regression, which indicates that the models do not over-fit the data. While the adjusted R^2 for Model 2 is about 2% greater than that of Model 1, the two quantities cannot be compared directly since the data sets

used for the two models are different. By averaging the data used in Model 2, the variance of the data for this model was decreased, therefore, Model 2 does not necessarily have greater explanatory power than Model 1. Furthermore, the predictive ability of Model 1 is better for the withheld data, as the predicted error sum of squares (PRESS) for the first model is 32% less than that of the second model. Moreover, the residuals for Model 1 follow the Normal distribution much more closely than those of Model 2.

Figure 5-8 – Histograms of Residuals, Models 1 and 2

Model 1 Rediduals

Std. Dev = .99
Mean = 0.00
N = 271.00

Regression Standardized Residual

Model 2 Residuals

Std. Dev = .99
Mean = 0.00
N = 271.00

Regression Standardized Residual
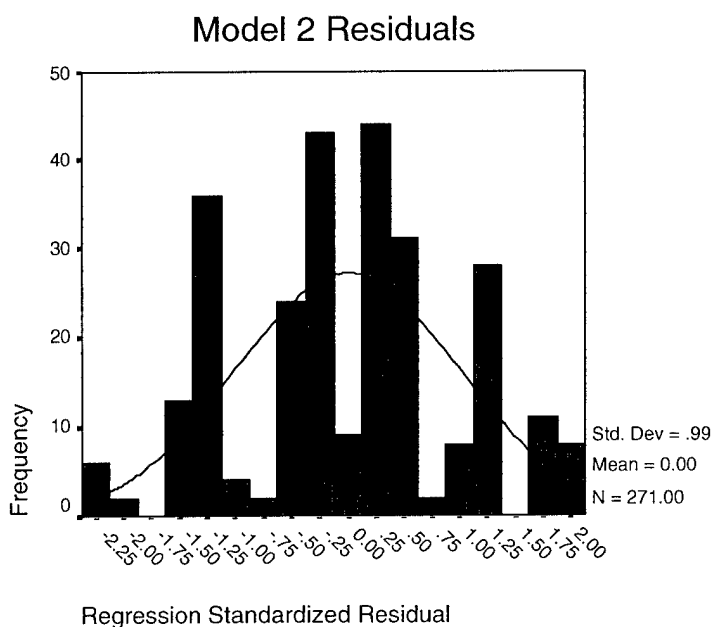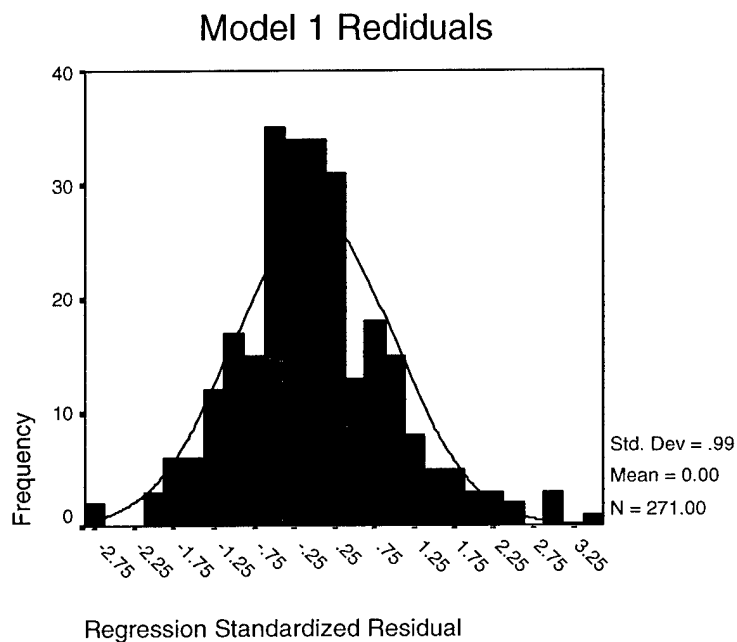
Figure 5-9 – Normal P-P Plots of Residuals, Models 1 and 2

Model 1

Observed Cum Prob

Model 2

Observed Cum Prob
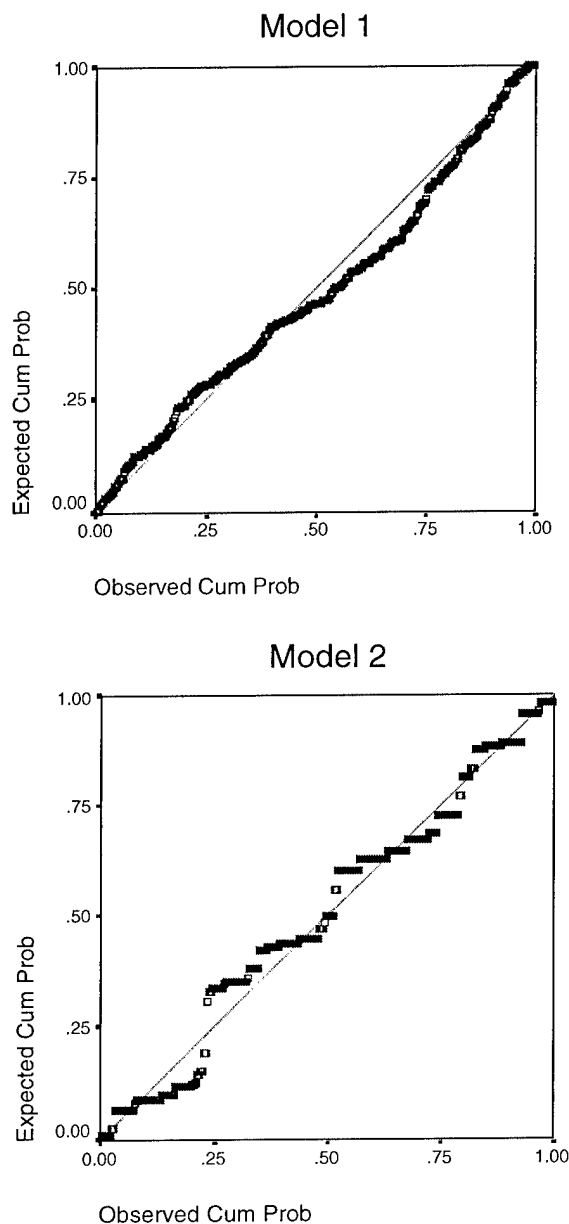
The better predictive ability of Model 1 and its better-behaved residuals suggest that it is the better of the two models.

The slope coefficients for the two models agree in sign and in magnitude (approximately). The sign of every coefficient is consistent with intuition. Manpower and Training Requirements, Corrective Maintenance, and Software Support and Maintenance Requirements all have positive slope coefficients with respect to O&S Cost while that of Ease of Upgrade/Technology Insertion is negative. Moreover, in Model 1, the binary variables agree with intuition. The variable for Regime (0 for pre-'94 and 1 for '94-'99) has a negative coefficient. Additionally, the variables for life cycle phase are all negative. Recall, that when all three of these are equal to zero, then

the system underwent a major upgrade for that year of data. Thus, when all three binary variables are zero, cost is higher (since the coefficients of all three variables are negative). Among the three binary variables, Phase 3, which corresponds to the Phase-Out portion of a system's life cycle, is the most negative. Phase 2, which corresponds to the Steady State portion, is the second most negative. Finally, Phase 1, which corresponds to the Build-Up portion, is the least negative.

The dominant direct cost driver in both regressions is System Manpower and Training Requirements. However, Corrective Maintenance has an additional indirect effect on cost in that it drives (partially) manpower requirements. (Recall, that the metric's leverage is the sum of its direct effect on cost and its indirect effect(s). This will be further discussed in the Chapter 6 where the final leverage results for each metric are presented.)

The metric for Ease of Upgrade/Technology Insertion does not meet conventional significance requirements of 0.05 (or 0.10) in either model, though the sign and magnitude of the coefficients for this metric roughly agree in both models. However, in Model 1, which is esteemed to be the better model, the coefficient for this metric has a p-value of 0.142. This is close to being significant at the 0.10-level. Naturally, there is a 14% chance that this coefficient actually is zero and the negative (i.e. cost-reducing) signal is due purely to chance. However, the coefficient's closeness to conventional levels of significance may at least suggest that the metrics for Ease of Upgrade/Technology Insertion are indicators of cost-saving constructs. One reason why these metrics did not turn out to be significant in the model is that cost savings from these concepts will only materialize if a system has been around long enough to be upgraded. There are many systems in the VAMOSC which have not been around long enough to be upgraded. Additionally, for many systems, the window of data in the VAMOSC reports may not go back far enough to capture upgrade costs. As more years of data become available, these metrics may prove significant in later analysis. Alternately, it may be that the metrics used to measure the degree to which a system lends itself to efficient, affordable upgrade are not valid measures for this construct.

The only point on which the two models really differ is the coefficient for Software Maintenance and Support Requirements. In Model 1, which is arguably the better model, the coefficient's p-value is only 0.128, whereas in Model 2, the p-value is 0.000 and the magnitude of the standardized coefficient is nearly triple that of Model 1. One reason why this coefficient is significant in Model 2 and not in Model 1 could be that by averaging the data for Model 2, the variance of the data was reduced. Therefore, the standard error of the slope coefficient was decreased by roughly 20%, making the coefficient's t-statistic significant in Model 2. Both models agree, however, in that the standardized coefficient for this metric is positive and smaller in absolute value than those of Manpower and Training Requirements and Corrective Maintenance, yet larger than that of Ease of System Upgrade/Technology Insertion.

Since Model 1 is arguably the better of the two models, the standardized coefficients from this model were used to calculate the final leverages reported in Chapter 6.

The scatter plot below shows the predicted vs. actual O&S Cost for Model 1 for the withheld data (from FY '98).

Figure 5-10 – Predicted Cost vs. Actual Cost for Withheld Data
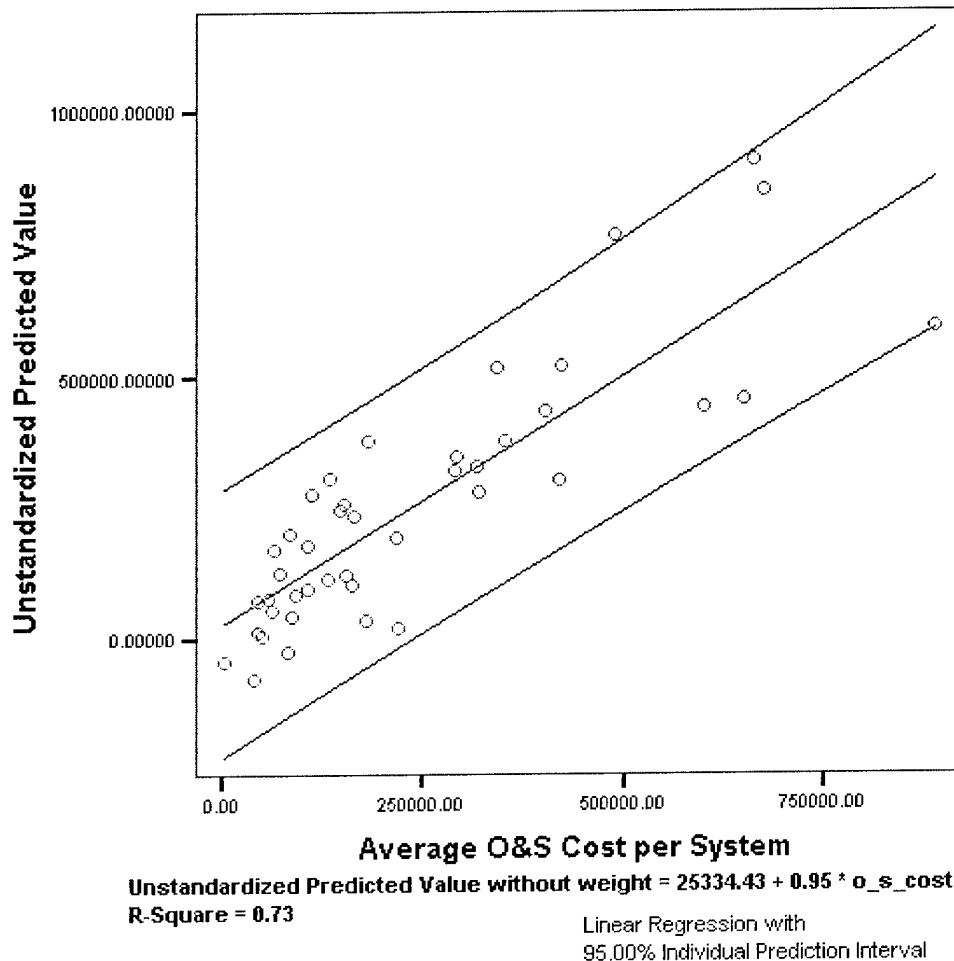
The plot reveals that for the most part, the model predicts accurately. Note, that the slope coefficient of the predicted values regressed on the actual values is 0.95 which is very close to 1.0, the ideal coefficient.

As the histograms and Normal P-P plots previously indicated, the residuals for Model 1 appear to follow the Normal distribution closely. On the other hand, the following plot suggests that while the variance of the residuals is approximately constant, the residuals drift upward with true O&S Cost. A log transformation of O&S Cost may eliminate this problem in future analysis. The current model tends to over-predict the cost of less expensive systems and under predict that of more expensive systems.

Figure 5-11 – Scatter Plot of Residuals vs. O&S Cost
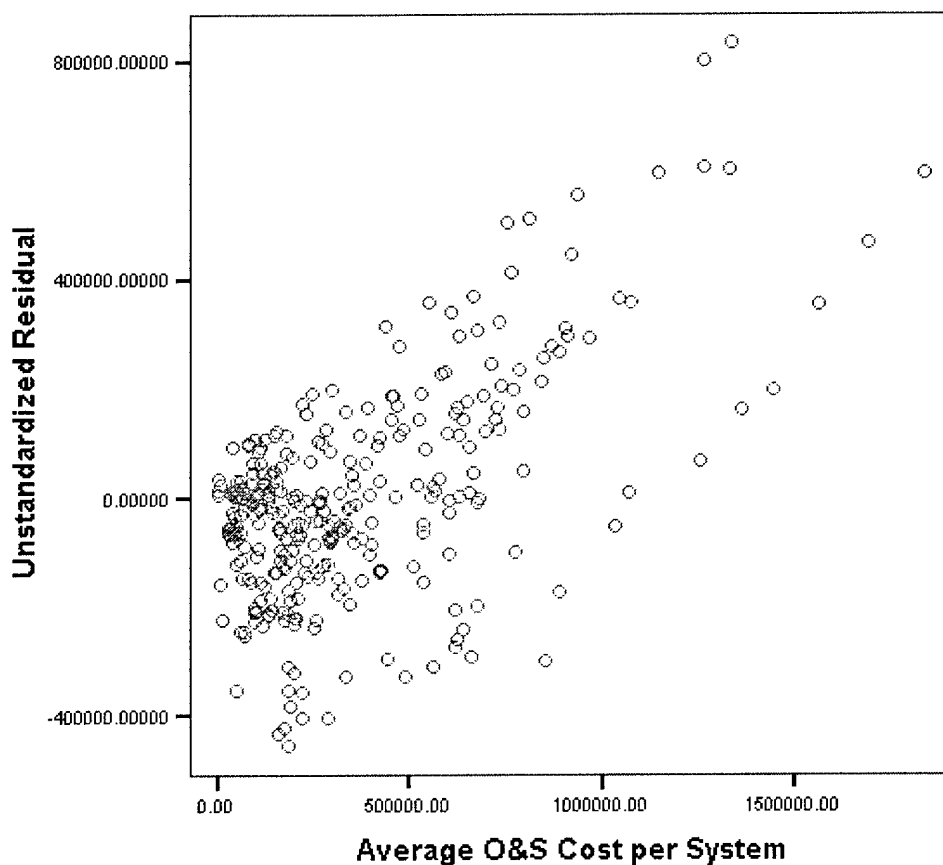
Influential Observations and Outliers:

As mentioned in Section 5-3, outliers and influential observations were flagged according to their leverage, influence, and the magnitude of their studentized residuals. In total, 11 of the observations that were identified by these criteria were excluded from the analysis (3.9% of the data points used in the regression, not including the withheld data). Only those observations for which something very unusual in the data was apparent were eliminated. Of the 11 excluded

data points, 4 were years during which there was a major modernization expenditure (such as the cost spike in Figure 5-3). Another 4 were excluded because the systems were at the very beginning of their life cycles and a low population size inflated their cost per system. The remaining 3 were excluded because they were at the very end of their life cycles and were no longer being funded or supported.

Mulitcollinearity:
Since System Manpower and Training Requirements was hypothesized to be a function of System Corrective Maintenance, it was likely that multicollinearity would pose a problem in the regression analyses. However, none of the variance inflation factors (VIF) associated with the variables exceed the conventional threshold of 10 (see Appendix 2). Additionally, none of the condition indices exceeded 15.

Since averaging the data did not yield an increase in explanatory power or predictive ability for this model, it was decided to use the non-averaged yearly data for the other regressions. Moreover, since System Manpower Requirements and System Training Requirements were merged into one metric, the regression of System Training Requirements on System Manpower Requirements and the Usability Metrics was not necessary.

5-4-2 Regression of System Manpower and Training Requirements

Three metrics were initially hypothesized to drive system Manpower and Training Requirements directly:

- Corrective Maintenance
- Degree of Automation
- Usability of the System

Corrective Maintenance:
The metrics for Corrective Maintenance were discussed in the previous sub-section.

Degree of Automation:
Only one metric was available for this construct, a survey question administered to representatives at the system program office, ISEA, and FTSCLANT technicians. (For the scale used, refer to Section 4-4-2). The reliability of this metric was relatively high (0.72) among respondents. The correlations among the three different respondents (for each system) were all highly significant.

Usability Metrics:

The metrics having to do with how well human operators and maintainers interface with the system were initially grouped under the heading "usability." These metrics were as follows:

- Sailor Proof-ness Survey Question (asked of program office, ISEA, and FTSCLANT technicians)
- The Number of Operator/Maintenance Induced Failures Reported per CSRR
- The Number of Inadequate Training Problems Reported per CSRR
- The Number of Inexperienced Personnel Problems Reported per CSRR

The reliability of these metrics was less than desired. Together, the three metrics pertaining to human related problems reported per CSRR exhibited a reliability of only 0.44. This relative lack of internal consistency could be due to the fact that these problems are underreported as some Navy analysts suggested. The internal reliability of the fourth metric, Sailor Proofness was somewhat higher, at 0.54. However, when summed, all four metrics exhibit very poor reliability, at 0.05. Clearly, all four metrics do not measure the same construct and could not be used together as metrics for usability.

A correlation matrix (see Appendix 3) indicates that Manpower and Training Requirements is highly correlated with the metrics for Corrective Maintenance and Degree of Automation. As expected, the correlation with Corrective Maintenance is positive, whereas the correlation with Degree of Automation is negative. Manpower and Training Requirements has significant, but weaker correlations the Number of Operator Induced Failures per CSRR and the Number of Inexperience Problems per CSRR. Although the three metrics pertaining to human related problems per CSRR exhibited low reliability, their sum (HMNPRBLM) has a significant positive correlation with Manpower and Training Requirements. The correlation with Sailor Proofness is negative (as expected), but very week and not significant.

Since the Usability metrics could not be grouped together, Manpower and Training Requirements was regressed on Corrective Maintenance, Degree of Automation, and the sum of human related problems (HMNPRBLM). The resulting regression yielded reasonable results, but little explanatory power. The metric for human related problems was excluded since it had a p-value of 0.660 when included in the regression. Furthermore, the predictive ability of the model without HMNPRBLM was better).

Table 5-8 – Regression Results for Manpower and Training

Explanatory Power (Data Set)	
R^2	0.249
Adjusted R^2	0.243
Significance of Model	0.000
Predictive Ability (FY 98 Data)	
Squared Correlation (R^2)	0.412
Standardized Coefficients with Significance	
Corrective Maintenance	+0.403 (.000)
Degree of Automation	-0.241 (.000)

Although the explanatory power of the model is somewhat lacking, the predictive ability of the model for the withheld data is better (R^2 of 0.412 vs. 0.249). The fact that the model can only explain about a quarter of the variance in Manpower and Training Requirements suggests that there are missing variables, not collected in the data set. The residuals from the regression suggest that the model systematically over-predicts Manpower and Training Requirements for systems with low requirements and under-predicts that of systems with high requirements.

Figure 5-12 – Scatter Plot of Residuals vs. Manpower & Training Requirements

Manpower and Training Requirements

The data points with the largest residuals all corresponded to a handful of systems with the greatest manpower requirements. Three of these systems are large sonar systems with consoles at which many human operators sit. Two others were variants of the same basic fire control system used with submarine sonar. The other system was a combat direction system which also has many consoles which must be attended by human operators. All of these systems have a relatively large number of consoles that must be attended by human operators. Unfortunately, no data was collected on the precise number of consoles that must be attended by human operators for each system. Judging by the residuals, it is likely that this information would help to improve the model.

5-4-3 Regression of System Corrective Maintenance

The metrics originally hypothesized to drive Corrective Maintenance were:

- Inherent Maintainability
- Inherent Reliability
- Usability

Inherent Maintainability:
As previously discussed, the metrics for usability did not coalesce well enough to be grouped together. Similar problems arose with the Maintainability metrics:

- Mean-Time-to-Repair (MTTR)
- CASREP Maintenance Hours/CASREP
- Corrective Maintenance Manhours/Corrective Maintenance Action
- Number of Technical Assist Visit Requests (TAVR) per system

Combined, these four metrics have a low reliability of 0.14. A correlation matrix with these four metrics and the dependent regression variable, Corrective Maintenance, also indicates that the metrics do not exhibit a high degree of consistency (see Appendix 4). MTTR is actually negatively correlated with the other three metrics, though the correlations are not significant. Moreover, MTTR has a negative correlation with Corrective Maintenance that has a significance of .058. Thus, MTTR consistently has correlations with the other metrics and Corrective Maintenance that are negative when one would expect them to be positive. Though none of these correlations is significant at the 0.05 level, the negative signs of the coefficients are puzzling. No explanation could be found for this. Since MTTR was not available for many of the data points and it exhibited insignificant correlations with the other metrics, it was excluded from the model. Likewise, CASREP Maintenance Hours/CASREP and Corrective Maintenance Manhours/Corrective Maintenance Action were excluded as they did not correlate strongly to any of the other metrics or the regression dependent variable. The lone metric to correlate strongly with Corrective Maintenance was TAVR/System. Note that TAVR/System is a negative indicator of system maintainability in that the higher this number, the lower (in theory) the maintainability of the system.

Inherent Reliability:
Mean-Time-Between-Failures was the only available measure of system reliability. The skewness of the metric's histogram and the non-linear shape of the scatter plot of MTBF and Corrective Maintenance suggested that a transformation of the metric was appropriate. The inverse transform and the natural logarithm transform were attempted and the latter provided the best results in terms of non-skewness and the correlation with Corrective Maintenance.

Figure 5-13 – Histograms of MTBF and LN(MTBF)

MTBF

Std. Dev = 1483.87
Mean = 1261.1
N = 156.00

MTBF

LOGMTBF

Std. Dev = .84
Mean = 6.75
N = 156.00
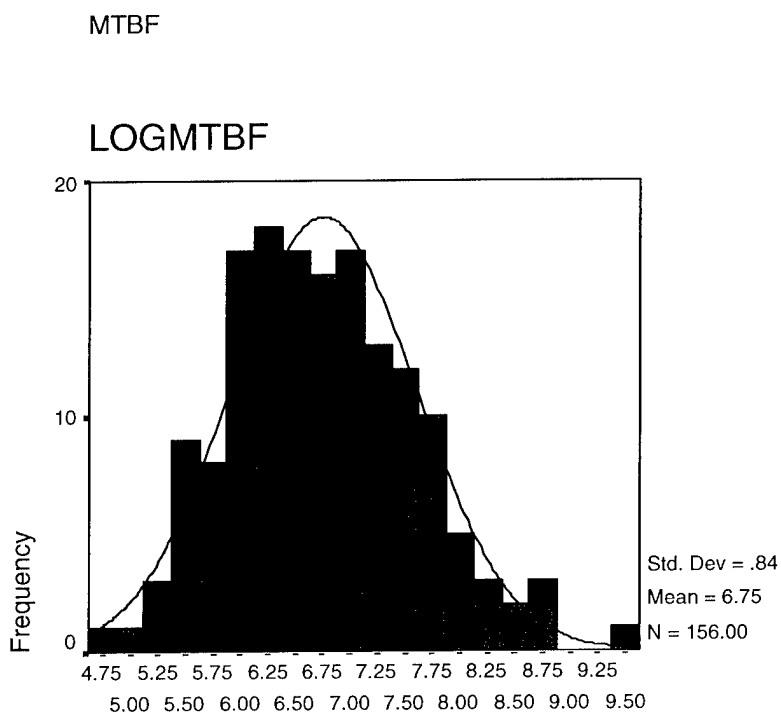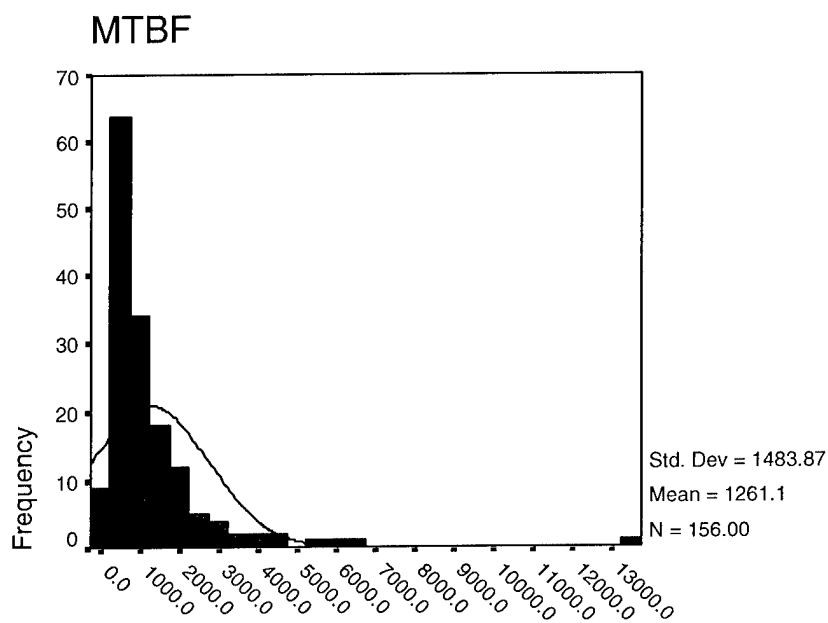
LOGMTBF

***Figure 5-14 – Scatter Plots of Corrective Maintenance vs. MTBF &
LN(MTBF)***
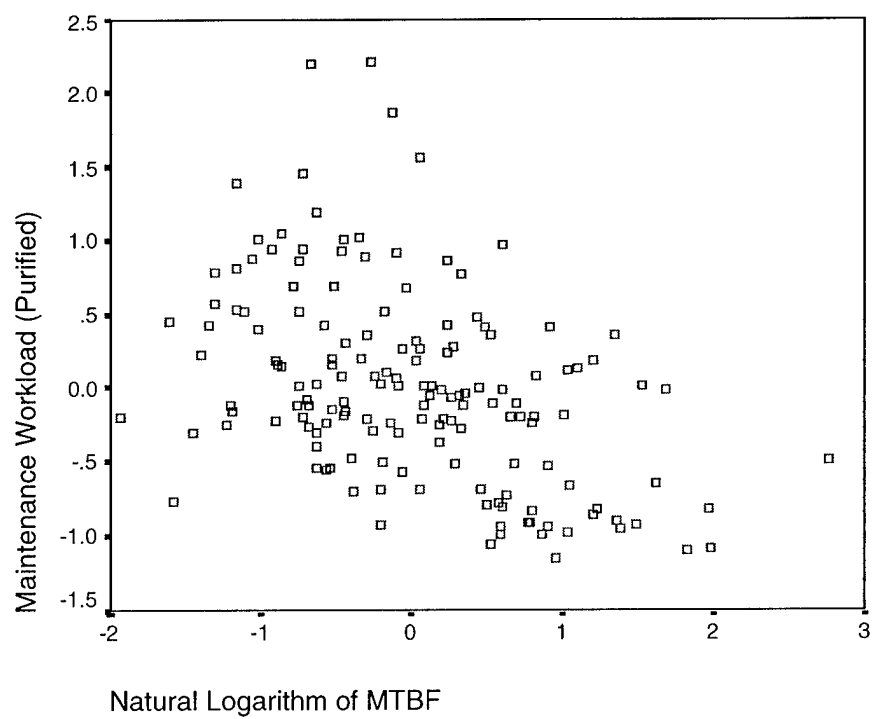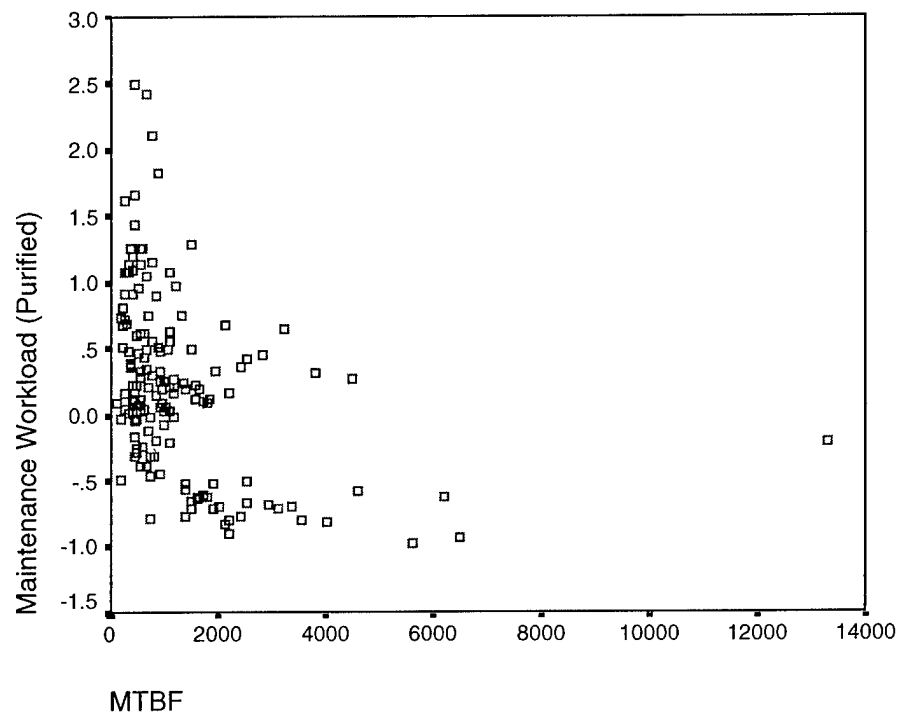
<u>Usability Metrics:</u>
Though they could not be grouped together, Sailor Proofness and the HMNPRBLM were considered separately as candidate variables for the regression of Corrective Maintenance.

Since the maintainability metrics did not exhibit a high degree of internal consistency, the metrics in the casual diagram that were hypothesized to drive Inherent Maintainability were moved up one level and made available as candidate variables for the regression of Corrective Maintenance. Therefore, Degree of Modularity and BIT/ATE Quality (A regression of TAVR on these variables was attempted without satisfactory results.) Therefore, the maintainability causal variables were moved up one level and made potential regression variables for Corrective Maintenance.

<u>BIT/ATE Quality:</u>
Two metrics were available as measures of the quality of a system's Built-In Testing or Automatic Test Equipment.

- BIT Quality Survey Question (asked of program office, ISEA, and FTSCLANT technicians)
- Number of BIT/ATE Problems Reported per CSRR

Unfortunately, these two metrics exhibited poor reliability, -0.13. Therefore, the two metrics could not be used together. However, the two metrics were both considered separately as candidate variables for the regression analysis.

<u>Modularity:</u>
This metric, also used as one of the Ease of Upgrade/Technology Insertion metrics, was hypothesized to affect the Inherent Maintainability of a system, and like the BIT/ATE Quality metrics, was moved up one level in the metrics hierarchy and made available for the regression of Corrective Maintenance.

In the final model, only three variables were significant, the natural logarithm of MTBF, TAVR/System, and Sailor Proofness. The following table summarizes the results:

Table 5-9 – Regression Results for Corrective Maintenance

Explanatory Power (Data Set)	
R^2	0.639
Adjusted R^2	0.615
Significance of Model	0.000
Predictive Ability (FY 98 Data)	
Squared Correlation (R^2)	0.507
Standardized Coefficients with Significance	
ln(MTBF)	-0.395 (.000)
TAVR/System	+0.467 (.000)
Sailor Proofness	-0.290 (.000)

The explanatory power of the model is much better than the previous model, and there is no great disparity between the R^2 for the data and the withheld data. This suggests that the model did not over-fit the data. Furthermore, the signs of the coefficients all agree with intuition.

It should be noted that because MTBF was available for so few systems, missing values for this analysis had to be excluded pair-wise as opposed to filling in missing values with mean values as in the previous regressions. The number of data points used in the regression was considerably lower with 49 data points (and 12 withheld FY '98 data points) with sufficient data to use in the analysis.

The following plots show the predicted values vs. actual values of Corrective Maintenance for the data set and the withheld data.

Figure 5-15 – Scatter Plot of Predicted vs. Actual Corrective Maintenance for Data Set

Maintenance Workload (Purified)

Unstandardized Predicted Value = 0.05 + 0.64 * maint
R-Square = 0.62

Linear Regression with
95.00% Individual Prediction Interval

91
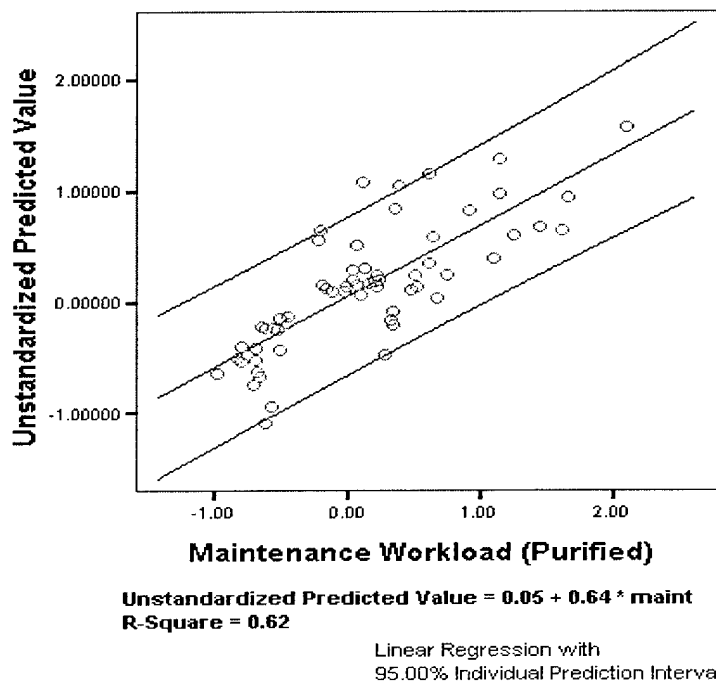
Figure 5-16 – Scatter Plot of Predicted vs. Actual Corrective Maintenance, Withheld Data

Maintenance Workload (Purified)
Unstandardized Predicted Value = 0.03 + 0.61 * maint
R-Square = 0.51

Linear Regression with
95.00% Individual Prediction Interval
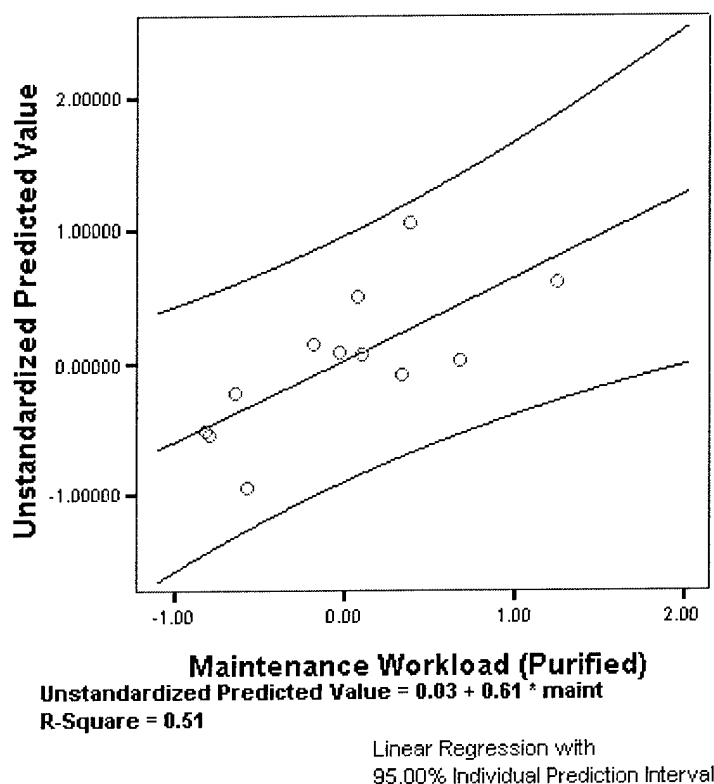
Though they significant in the regression (and were, therefore excluded from the model), it is worthwhile to note that the BIT Quality Survey Question and the number of human related problems per CSRR correlate significantly with Corrective Maintenance:

Table 5-10 – BIT Quality & Human Related Problems, Correlations with Corrective Maintenance

Metric	Correlation to Corrective Maintenance	Significance of Correlation
BIT Quality	-.150	.007
HMNPRBLMS	+.254	.001

92

6-1 *Tabulation of Leverages*

The table below presents the final leverage estimates for the strategic metrics that directly drive O&S Cost.

Table 6-1 – Cost Leverages for Strategic Metrics

Strategic Metric	Leverage	Significance in Cost Regression
Manpower & Training Requirements	+0.564	.000
Corrective Maintenance	+0.460	.000
Software Support & Maintenance	+0.068	.128
Upgradability of System	-0.055	.142

Although Manpower and Training Requirements has the highest estimated leverage on O&S Cost, the leverages in Table 6-1 are estimates, and therefore, contain some error. It should, therefore, be noted that the leverages of Manpower and Training Requirements and Corrective Maintenance are very close in magnitude, statistically. On the other hand, since the manpower costs in the VAMOSC reports do not include the cost of officers and un-assigned personnel who operate and maintain the systems in the reports, it is likely that this is an underestimate of the true cost. Furthermore, the VAMOSC reports do not include the less visible personnel costs of retirement benefits or benefits accorded to spouses and dependents of the service members who operate and maintain the systems.

As for the other metrics in the table, they do not meet the conventional standards for significance, yet they are close to being significant at the 0.10-level and are, therefore, included in the table as potential indicators of cost affecting constructs.

The table below presents the final leverages (with respect to O&S Cost) estimated for the subordinate causal metrics for the strategic cost driving metrics.

Table 6-2 – Cost Leverages for Subordinate Metrics

Subordinate Metric	Strategic Metric	Leverage with Respect to O&S Cost
Degree of Automation	Manpower & Training	-0.136
Reliability: ln(MTBF)	Corrective Maintenance	-0.182
Maintainability: TAVR/System*	Corrective Maintenance	+0.215
Sailor Proofness	Corrective Maintenance	-0.133

* TAVR/System, recall, is a negative indicator of Maintainability, or alternately, an indicator of Un-Maintainability.

Though TAVR/System is an indicator of the maintainability of a system, it is also a partial indicator of reliability in that for a TAVR to occur, a failure must first occur. Though collinearity diagnostics did not reveal a problematic closeness of association between these metrics from the perspective of regression analysis, the two metrics do have a significant correlation.

Table 6-3 – Correlations of TAVR/System and Reliability

	Correlation with TAVR/System (Significance)
MTBF	-0.210 (.088)
ln(MTBF)	-0.283 (.021)

Therefore, the true leverage of Inherent Reliability is likely greater than the leverage estimated for ln(MTBF). By the same token, the leverage of Inherent Maintainability is likely less than the estimated leverage for TAVR/System since some of this metric's leverage may in truth, be attributable to Reliability.

Though it is necessary to measure the risk discount factors (RDF) to calculate the incentive weights for these metrics using Equation 3.4, time did not allow for the measurement of this term. Therefore, the incentive weights of these metrics could not be calculated. However, in previous applications of the Metrics Thermostat, the leverage term dominated Equation 3.4 and the relative prioritization of metrics as determined by Equation 3.4 (using the RDF) did not differ from the relative prioritization suggested by the leverages. In other words, it is very likely that when the RDF are measured, the weights calculated for each metric using Equation 3.4 will not differ significantly from the leverages presented here.

A revised causal diagram is shown on the next page. The thickness of the arcs represents the magnitude of the corresponding leverages and asterisks denote negative coefficients on the corresponding arcs.

Figure 6-1 Revised Causal Diagram: Hierarchy of Metrics
*Denotes a negative causal relationship

Software Support and Maintenance Requirements
1. Size of Software (KSLOC)
2. Complexity of Code
3. Robustness of Software to Hardware Changes

O&S Cost/System

Degree of Automation*
1. Survey Question

Manpower & Training Requirements
1. NEC's Assigned/System-
2. Personnel to Operate & Maintain/System (according to Program Office)
3. Student Training Days/System

Ease of System Upgrade/Technology Insertion*
6. Upgradability Survey Question
7. Use of Open Architecture/Open Standards
8. Use of COTS Components
9. Modularity
10. Use of Non Development Items

Usability*
4. Sailor Proof-ness Survey Question

System Corrective Maintenance
6. Maintenance Workload Survey Question
7. Corrective Maintenance Actions/System
8. Corrective Maintenance Manhours/System
9. CASREP Maintenance Manhours/System

Inherent Reliability*
1. ln[MTBF]

Inherent Maintainability/Reliability
1. TAVR/System

95

-This page intentionally left blank-

The results of this analysis are preliminary and there are many ways in which to build on these first results.

First, the existing data set must be filled out more. Filling in missing values with data as opposed to mean values will likely strengthen the significances of some of the metrics for which much data is missing, particularly, the metrics pertaining to Upgradability and Software Support and Maintenance Requirements. In addition to filling in missing data for the metrics that are already in the model, adding new metrics to the data set may also increase the explanatory power of the models used to estimate the leverages. For example, measuring the number of operator consoles each system has may help improve the regression model for Manpower and Training Requirements. Substituting the Mean-Time-Between-Corrective-Maintenance for MTBF may also improve the explanatory power and predictive ability of the regression of Corrective Maintenance since MTBF only tracks critical failures that take the system down and not other, less critical failures that still require corrective maintenance. Including some measure of utilization or stress time for the systems may further improve the analysis. Steaming hours are available for many systems and stress time is available for those systems for which the MRDB maintains data (unfortunately, there was not enough time to obtain them for this research). In addition, accurate measures of how much preventative maintenance is actually performed on the systems may also improve the models.

The second major area for improvement is that of the reliability of the metrics, particularly those in the form of survey questionnaires. Of the 12 survey questions, 3 had a Cronbach's alpha less than 0.5, and 8 had reliability less than 0.6. Reducing the ambiguity of the scales used to rate the systems may prove beneficial. Having more respondents rate each system or having a few individuals familiar with all the systems rate all systems may also have this effect, though it was difficult enough getting 2 or 3 people to rate each system and none of the people interviewed felt qualified to rate all of the systems in the data set.

One of the working assumptions of the Metrics Thermostat (Step 1) is that all of the data points are sufficiently homogeneous. As previous implementations of the Metrics Thermostat had suffered from a small number of data points, at the outset of this research, data was collected for as many systems as possible. This eagerness to collect as much data as possible may have resulted in a data set with systems too diverse to be analyzed together. Narrowing the focus of the research on a group of more similar systems, such as radars or sonars, may result in better (more linear and less variance) data for regression analyses. Additionally, this might make it possible to have each person responding to the survey questions rate all systems in the study. The measurement scales might be more finely tuned to the type of system selected for analysis also.

Thus far, only Steps 1-4 of the Metrics Thermostat have been attempted with Navy systems. The remaining three are yet to be attempted.

Step 5 entails using survey measures to obtain the Risk Discount Factor (RDF_i) for each metric. Those taking the surveys in this context would be those who develop new systems, most likely defense contractors like Raytheon or General Dynamics.

In Step 6, Equation 3.4 is used to calculate ($\hat{w}^d_i$) for each metric. Emphasis on each metric is increased or decreased as indicated by Equation 3.4. Though measurements of the RDF_i are necessary to calculate the weight of each metric, in previous applications of the Metrics Thermostat, the leverage term in Equation 3.4 has dominated the RDF term. Therefore, it is likely that the relative weights of the metrics in this research would not differ substantially from their respective leverages.

Finally, Step 7 mandates periodic returns to Step 3 to update ($\hat{w}^d_i$). The timing of the periodic returns to Step 3 should correspond to significant changes in the operating environment, such as the emergence of new, revolutionary technologies, or new priorities at NAVSEA (i.e. when a new set of buzz words like "COTS" or "faster, better, cheaper," or "quality of life" emerges, indicating a shift in Navy procurement culture).

Once the research is ready to proceed to these steps, representatives from the Navy must work to structure contract incentives such that the incentives based on the ($\hat{w}^d_i$) are not "gamed" by contractors (like the examples at Bausch & Lomb or H.J. Heinz, sited in Chapter 3) and that new products actually cost less to operate and support.

In closing, it is perhaps most appropriate to remark of this preliminary research is that it demonstrates the feasibility (as well as some of the difficulties) in applying the Metrics Thermostat to Navy acquisitions.

BIBLIOGRAPHY

1. Baker, George (1992), "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 100, 598-614.

2. Blanchard, Benjamin S (1998). *Logistics Engineering and Management.* Prentice Hall, Upper Saddle River, New Jersey.

3. Cronbach, L. J. (1951). "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika*, 16:3, 297-334.

4. DSMC (1997). *Acquisitions Logistics Guide.* Defense Systems Management College, Fort Belvoir, VA, December.

5. Frank, Carl (2000). "Metrics Thermostat for Strategic Priorities in Military System Acquisition Projects," S.M. thesis, Management of Technology, MIT, Cambridge, MA (June).

6. Gibbons, Robert (1997), "Incentives and Careers in Organizations," in D. Kreps and K. Wallis (Eds.) *Advances in Economic Theory and Econometrics*, Cambridge, UK: Cambridge University Press, 1-37.

7. Hauser, John R. and Clausing (1988). "The House of Quality," *Harvard Business Review*, 1, 63-73.

8. Hauser, John R. and Gerald Katz (1998). "Metrics: You Are What You Measure!" *European Management Journal*, 16, 5, 516-528.

9. Hauser, John R. (2000). "Metrics Thermostat," Massachusetts Institute of Technology, Sloan School of Management Working Paper 1-28.

10. Kamisnski, Paul R. (1995). "Reducing Life Cycle Costs for New and Fielded Systems," memorandum issued December 4, 1995. (CAIV Working Group paper, Attachment 2 also referenced).

11. LaFountain, Burt D. (1999). "An Empirical Exploration of Metrics for Product Development Teams," S.M. thesis, Mechanical Engineering Department, MIT, Cambridge, MA (January).

12. Majumder, Arpita (2000). "Strategic Metrics for Product Development at Ford Motor Company," S.M. thesis, Engineering and Management, MIT, Cambridge, MA (June).

13. Maremount, Mark (1995). "Blind Ambition: How the Pursuit of Results Got Out of Hand at Bausch & Lomb." *Business Week,* Oct. 23.

14. Naval Post Graduate School (NPS). PowerPoint presentation by CDR Jeff Cuskey, August, 1999. .

15. Office of the Secretary of Defense Cost Analysis Improvement Group (CAIG). "Operating and Support Cost Estimating Guide." May 1992.

16. Perry, (1994). *The Problem – Why Change is Necessary,* February, 1994.

17. Perry (1996). Prepared remarks by SecDef William J. Perry to the Military Academy of the Russian General Staff, Moscow, Oct. 17, 1996 *Defense Issues*, Vol. 11, Nr. 97, "A Pragmatic U.S.-Russian Partnership."

18. Peter, J. Paul (1979). "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," *Journal of Marketing Research,* 16 (February), 6-17.

19. PHD NSCC (1999). "PHD NSWC Safety, Effectiveness, and Affordability Review Guide." Unpublished paper.

20. Post, Richard J. and Kenneth E. Goodpaster (1981). "H. J. Heinz Company: The Administration of Policy," Harvard Business School Case #382-034.

21. R-TOC Working Group (2000). "Reduction of Total Ownershp Costs (R-TOC): Recent History and Future Prospects." Unpublished Draft.

22. Russell (2000). "Reengineering Metrics Systems for Aircraft Sustainment Teams: A Metrics Thermostat for Use in Strategic Priority Managment," S.M. thesis, Department of Aeronautics and Astronautics, MIT, Cambridge, MA (December).

23. VAMOSC. "Data Reference Manual for Shipboard Systems" available at www.ncca.navy.mil/vamosc.

24. Welsch, Roy (2000). 15.077 Modern Regression and Multivariate Analysis Course Reader, Spring 2000, MIT.

Appendix 1: Detailed Measures of Metrics
(Numbers in parentheses are Cronbach's α.)

Strategic Metrics:

Manpower and Training Requirements (0.65)
Constituent Metrics:
- Personnel Assigned NEC's per System
- Personnel per System According to the Program Office
- Student Training Days per System

Corrective Maintenance (0.71)
Constituent Metrics:
- Maintenance Workload Survey Question (0.81)
- Number of Corrective Maintenance Actions per System per Year
- Number of CASREP Man-hours per System per Year

Software Support and Maintenance Requirements (0.59)
Constituent Metrics:
- Thousands of Lines of Source Code (0.95)
- Complexity of Code Survey Question (0.79)
- Robustness of Code to Hardware Changes (0.50)

Ease of System Upgrade/Technology Insertion (0.73)
Constituent Metrics:
- Upgradability Survey Question (0.42)
- Use of Open Architecture/Open Standards Survey Question (0.55)
- Use of Commercial Components Survey Question (0.41)
- Modularity Survey Question (0.59)

Subordinate Metrics:
Degree of Automation Survey Question (0.72)
Technical Assist Visit Requests/System
MTBF (natural logarithm transformation)
Sailor Proofness Survey Question (0.54)

Appendix 2: Selected SPSS Output for Regression of Operating and Support Cost

Correlations of Corrective Maintenance Variables and O&S Cost

Correlations

		O&S Cost/System	CASREPS per System	CASREP Maintenance Hours per System	CM Actions per system	CM hours per system	Maintenance Workload Survey Question
O&S Cost/System	Pearson Correlation	1.000	.029	.353**	.418**	.366**	-.234**
	Sig. (2-tailed)	.	.662	.000	.000	.000	.000
	N	378	223	222	367	367	378
CASREPS per System	Pearson Correlation	.029	1.000	.023	.105	.086	-.117
	Sig. (2-tailed)	.662	.	.733	.127	.210	.082
	N	223	223	222	214	214	223
CASREP Maintenance Hours per System	Pearson Correlation	.353**	.023	1.000	.352**	.404**	-.143*
	Sig. (2-tailed)	.000	.733	.	.000	.000	.033
	N	222	222	222	213	213	222
CM Actions per system	Pearson Correlation	.418**	.105	.352**	1.000	.793**	-.220**
	Sig. (2-tailed)	.000	.127	.000	.	.000	.000
	N	367	214	213	367	366	367
CM hours per system	Pearson Correlation	.366**	.086	.404**	.793**	1.000	-.135**
	Sig. (2-tailed)	.000	.210	.000	.000	.	.000
	N	367	214	213	366	367	367
Maintenance Workload Survey Question	Pearson Correlation	-.234**	-.117	-.143*	-.220**	-.185**	1.000
	Sig. (2-tailed)	.000	.082	.033	.000	.000	.
	N	378	223	222	367	367	378

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

103

Regression Results for Operating and Support Cost

Model Summary[b,c]

Model	R		R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
	INCLUDE = 1.00 (Selected)	INCLUDE ~= 1.00 (Unselected)				R Square Change	F Change	df1	df2	Sig. F Change
1	.826[a]	.629	.682	.673	184128.7	.682	70.315	8	262	.000

a. Predictors: (Constant), Software Support and Maintenance Requirements, REGIME, Upgradability, PHASE_2, PHASE_1, Maintenance Workload (Purified), Manpower and Training Requirements, PHASE_3

b. Unless noted otherwise, statistics are based only on cases for which INCLUDE = 1.00.

c. Dependent Variable: Average O&S cost on a per system basis (without ammo & gas)

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.9E+13	8	2.4E+12	70.315	.000[a]
	Residual	8.9E+12	262	3.4E+10		
	Total	2.8E+13	270			

a. Predictors: (Constant), Software Support and Maintenance Requirements, REGIME, Upgradability, PHASE_2, PHASE_1, Maintenance Workload (Purified), Manpower and Training Requirements, PHASE_3

b. Dependent Variable: Average O&S cost on a per system basis (without ammo & gas)

c. Selecting only cases for which INCLUDE = 1.00

Coefficients[a,b]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	630135.6	50862.052		12.389	.000	529985.1	730286.0		
PHASE_1	-142419	33809.708	-.200	-4.212	.000	-208993	-75846.0	.538	1.859
PHASE_2	-215244	34738.055	-.293	-6.196	.000	-283646	-146843	.541	1.848
PHASE_3	-289916	35425.034	-.399	-8.184	.000	-359670	-220162	.511	1.957
Manpower and Training Requirements	201041.1	16310.174	.564	12.326	.000	168925.4	233156.8	.579	1.729
Maintenance Workload (Purified)	99386.397	17596.246	.233	5.648	.000	64738.336	134034.5	.710	1.408
REGIME	-76935.3	23890.590	-.119	-3.220	.001	-123977	-29893.3	.887	1.128
Upgradability	-27209.4	18493.662	-.055	-1.471	.142	-63624.5	9205.759	.860	1.163
Software Support and Maintenance Requirements	28669.167	18759.136	.068	1.528	.128	-8268.694	65607.027	.618	1.617

a. Dependent Variable: Average O&S cost on a per system basis (without ammo & gas)

b. Selecting only cases for which INCLUDE = 1.00

105

Collinearity Diagnostics[a,b]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions								
				(Constant)	PHASE_1	PHASE_2	PHASE_3	Manpower and Training Requirements	Maintenance Workload (Purified)	REGIME	Upgradability	Software Support and Maintenance Requirements
1	1	3.479	1.000	.00	.01	.01	.01	.00	.00	.02	.00	.00
	2	2.006	1.317	.00	.00	.02	.00	.09	.09	.00	.00	.09
	3	1.063	1.809	.00	.18	.00	.13	.00	.02	.01	.00	.02
	4	.932	1.932	.00	.06	.19	.06	.08	.02	.00	.00	.06
	5	.639	2.332	.00	.01	.07	.04	.01	.68	.00	.00	.17
	6	.394	2.972	.00	.01	.02	.00	.76	.09	.01	.00	.61
	7	.316	3.318	.01	.00	.08	.10	.01	.00	.93	.01	.00
	8	.142	4.947	.03	.65	.49	.50	.01	.08	.02	.12	.00
	9	2.881E-02	10.989	.96	.08	.11	.15	.03	.02	.01	.86	.04

a. Dependent Variable: Average O&S cost on a per system basis (without ammo & gas)

b. Selecting only cases for which INCLUDE = 1.00

106

Casewise Diagnostics[b,c]

Case Number	Status	Std. Residual	Average O&S cost on a per system basis (without ammo & gas)	Predicted Value	Residual
15	X[a]	4.272	1340819	554307.35	786511.6
16	X[a]	6.794	1803758	552818.73	1250939
49	X[a]	4.097	1268874	514496.67	754377.3
66	X[a]	7.137	1802724	488582.07	1314142
68		3.071	1843683	1278138.6	565544.7
90		3.080	1264416	697313.99	567102.0
94		3.054	1146612	584259.73	562352.3
103	X[a]	4.354	2154876	1353124.0	801751.8
126	X[a]	4.990	1600936	682046.81	918889.2
158	X[a]	3.238	936636.0	340469.50	596166.5
171		3.413	1336279	707845.72	628433.3
340	X[a]	3.923	976056.0	253775.80	722280.2
346	X[a]	3.211	1166705	575435.65	591269.4
347	X[a]	4.421	1271107	457121.01	813986.0

a. INCLUDE ~= 1.00 (Unselected)

b. Dependent Variable: Average O&S cost on a per system basis (without ammo & gas)

c. When values are missing, the substituted mean has been used in the statistical computation.

107

Residuals Statistics[a,b]

	INCLUDE = 1.00 (Selected)					INCLUDE ~= 1.00 (Unselected)				
	Minimum	Maximum	Mean	Std. Deviation	N	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-31439.3	1278139	360910.0	265770.9	271	-79294.4	1353124	331752.5	241629.6	107
Residual	-491474	628433.3	-8.01E-11	181380.4	271	-515535	1314142	34826.74	319512.9	107
Std. Predicted Value	-1.476	3.451	.000	1.000	271	-1.656	3.733	-.110	.909	107
Std. Residual	-2.669	3.413	.000	.985	271	-2.800	7.137	.189	1.735	107

a. Dependent Variable: Average O&S cost on a per system basis (without ammo & gas)

b. Pooled Cases

Charts

Model 1 Redividuals
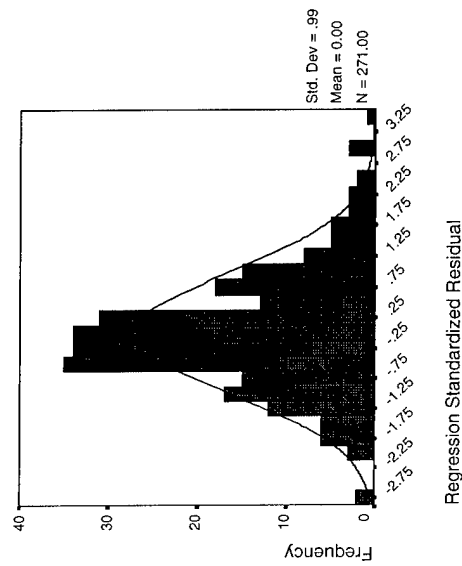
Std. Dev = .99
Mean = 0.00
N = 271.00

Regression Standardized Residual

108

Model 1

Partial Regression Plot

□ Unselected Cases

□ Selected Cases

Partial Regression Plot

Partial Regression Plot

110

Appendix 3: Selected SPSS Output for Regression of Manpower and Training Requirements
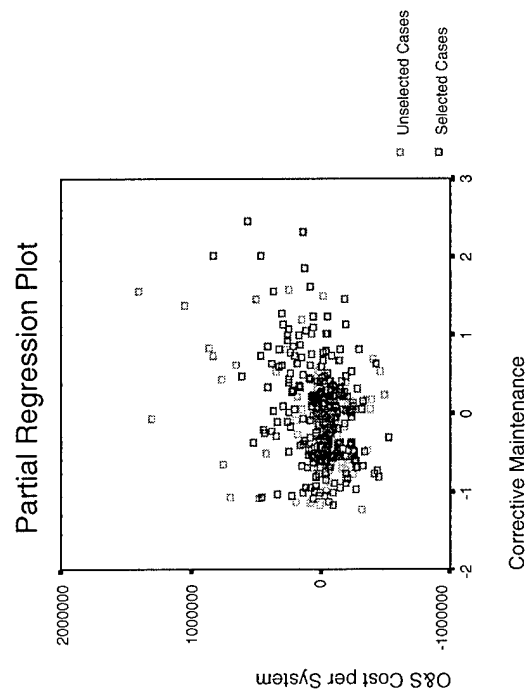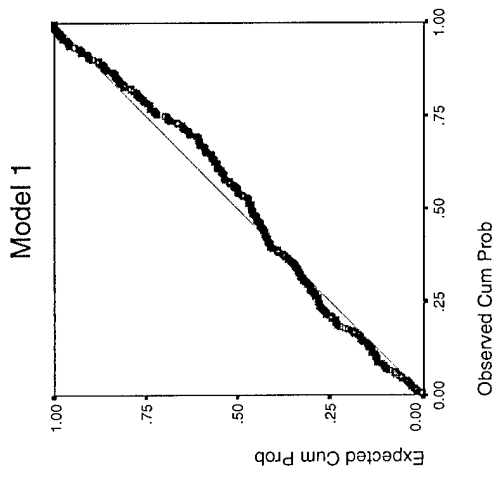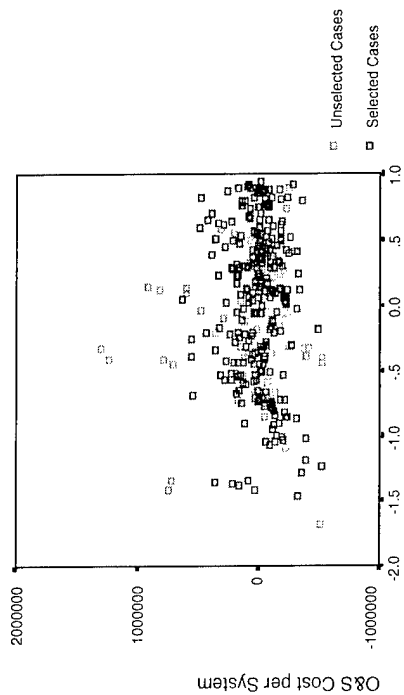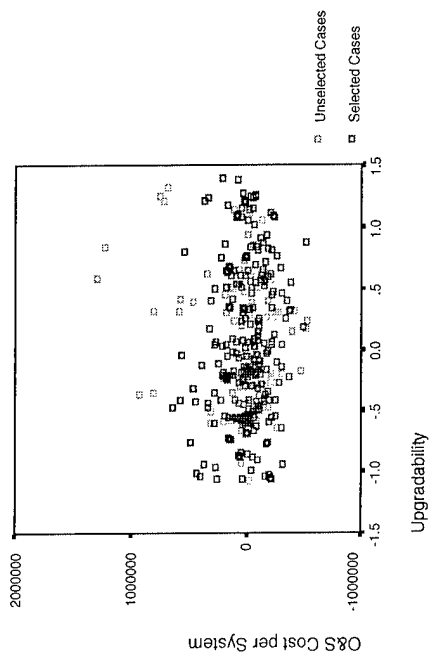
Correlation Matrix for Manpower and Training Requirements and Causal Variables

Correlations

	Manpower and Training Requirements	Sailor Proofness	Operator Induced Failures per CSRR	Inadequate Training Problems per CSRR	Inexperience Problems per CSRR	Maintenance Workload (Purified)	Automation	HMNPRBLM
Manpower and Training Requirements	1.000 / .	-.049 / .345	.177* / .013	.111 / .118	.159* / .025	.454** / .000	-.351** / .000	.220** / .002
Sailor Proofness	-.049 / .345	1.000 / .	.054 / .448	-.055 / .442	.030 / .673	-.245** / .000	.363** / .000	.028 / .695
Operator Induced Failures per CSRR	.177* / .013	.054 / .448	1.000 / .	.178* / .012	.217** / .002	.198** / .005	-.051 / .481	.733** / .000
Inadequate Training Problems per CSRR	.111 / .118	-.055 / .442	.178* / .012	1.000 / .	.292** / .000	.158* / .026	-.161* / .024	.602** / .000
Inexperience Problems per CSRR	.159* / .025	.030 / .673	.217** / .002	.292** / .000	1.000 / .	.143* / .045	-.095 / .185	.676** / .000
Maintenance Workload	.454** / .000	-.245** / .000	.198** / .005	.158* / .026	.143* / .045	1.000 / .	-.188** / .000	.243** / .001
Automation	-.351** / .000	.363** / .000	-.051 / .481	-.161* / .024	-.095 / .185	-.188** / .000	1.000 / .	-.133 / .063
HMNPRBLM	.220** / .002	.028 / .695	.783** / .000	.602** / .000	.676** / .000	.243** / .001	-.133 / .063	1.000 / .

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Regression Results for Operating and Support Cost

Model Summary[b,c]

Model	R				
			R Square	Adjusted R Square	Std. Error of the Estimate
	INCLUDE2 = 1.00 (Selected)	INCLUDE2 ~= 1.00 (Unselected)			
1	.499[a]	.642	.249	.243	.8103

a. Predictors: (Constant), Automation (averaged), Maintenance Workload (Purified)

b. Unless noted otherwise, statistics are based only on cases for which INCLUDE2 = 1.00.

c. Dependent Variable: Manpower and Training Requirements

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	59.353	2	29.677	45.198	.000[a]
	Residual	179.251	273	.657		
	Total	238.604	275			

a. Predictors: (Constant), Automation (averaged), Maintenance Workload (Purified)

b. Dependent Variable: Manpower and Training Requirements

c. Selecting only cases for which INCLUDE2 = 1.00

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.607	.137		4.441	.000
	Maintenance Workload (Purified)	.511	.067	.403	7.599	.000
	Automation (averaged)	-.201	.044	-.241	-4.540	.000

a. Dependent Variable: Manpower and Training Requirements

b. Selecting only cases for which INCLUDE2 = 1.00

Casewise Diagnostics[a]

Case Number	Std. Residual	Manpower and Training Requirements	Predicted Value	Residual
19	3.205	2.95	.3498	2.5968
20	3.393	3.07	.3180	2.7495
21	3.760	3.32	.2744	3.0466
22	3.595	3.17	.2617	2.9128
23	3.793	3.35	.2782	3.0737
24	3.660	3.24	.2786	2.9658
25	4.156	3.69	.3183	3.3677

a. Dependent Variable: Manpower and Training Requirements

Residuals Statistics[a,b]

	INCLUDE2 = 1.00 (Selected)					INCLUDE2 ~= 1.00 (Unselected)				
	Minimum	Maximum	Mean	Std. Deviation	N	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-1.1281	1.3428	4.974E-02	.4646	276	-1.0478	1.3752	1.579E-02	.5226	102
Residual	-1.3375	3.3677	-2.69E-16	.8074	276	-1.4623	1.8224	-8.81E-02	.5615	102
Std. Predicted Value	-2.535	2.783	.000	1.000	276	-2.362	2.853	-.073	1.125	102
Std. Residual	-1.651	4.156	.000	.996	276	-1.805	2.249	-.109	.693	102

a. Dependent Variable: Manpower and Training Requirements

b. Pooled Cases

Charts
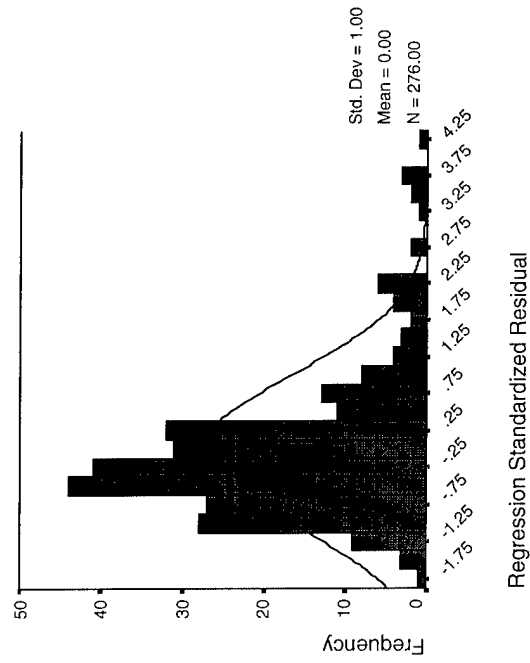
Histogram of Residuals

Std. Dev = 1.00
Mean = 0.00
N = 276.00

Regression Standardized Residual

114

Normal P-P Plot of Standardized Residuals

Expected Cum Prob

Observed Cum Prob

Scatterplot

Regression Standardized Residual

Manpower and Training Requirements

Unselected Cases

Selected Cases

115

Partial Regression Plot

Maintenance Workload (Purified)

Manpower and Training Requirements

□ Unselected Cases

Selected Cases

Partial Regression Plot

Automation (averaged)

Manpower and Training Requirements

□ Unselected Cases

Selected Cases

116
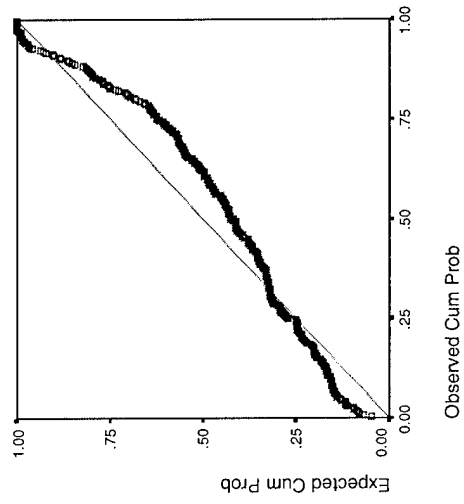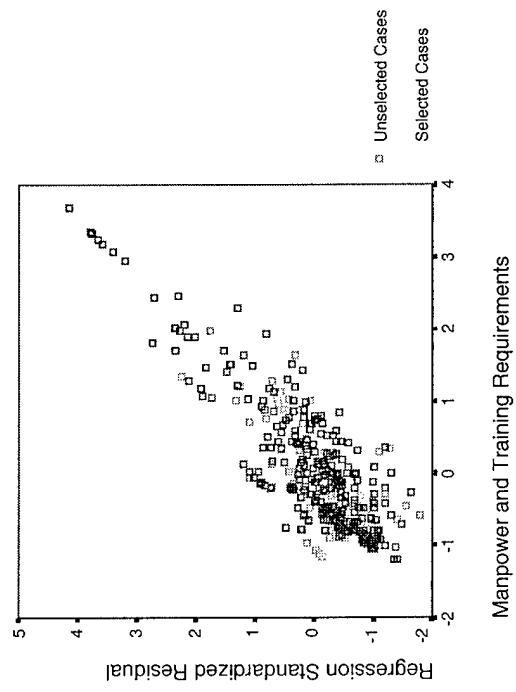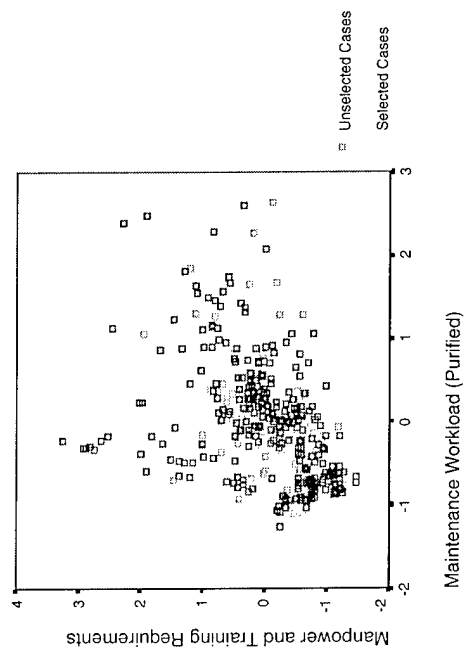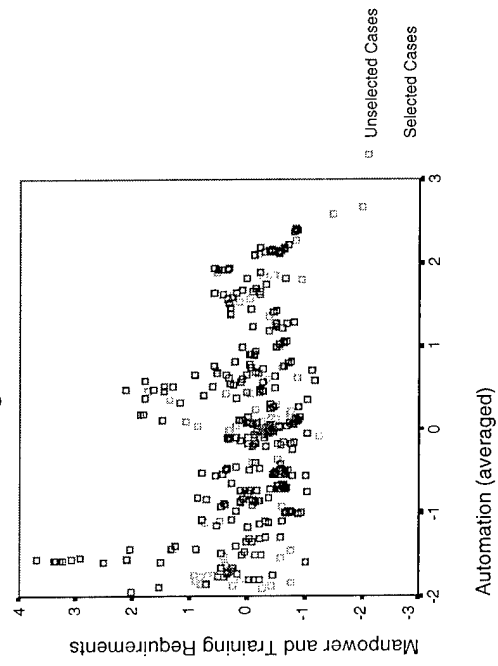
Appendix 4: SPSS Output for Regression of Corrective Maintenance

Correlations of Corrective Maintenance and Maintainability Metrics

Correlations

		CASREP Maintenance Hours per CASREP	Tech Assist Visit Requests per System	CM Time per CM Action	MTTR	Corrective Maintenance (Purified)
CASREP Maintenance Hours per CASREP	Pearson Correlation	1.000	.145	.076	-.070	.102
	Sig. (2-tailed)	.	.078	.267	.491	.127
	N	223	149	213	100	223
Tech Assist Visit Requests per System	Pearson Correlation	.145	1.000	.092	-.195	.455**
	Sig. (2-tailed)	.078	.	.269	.108	.000
	N	149	155	146	69	155
CM Time per CM Action	Pearson Correlation	.076	.092	1.000	-.018	.048
	Sig. (2-tailed)	.267	.269	.	.821	.361
	N	213	146	360	162	360
MTTR	Pearson Correlation	-.070	-.195	-.018	1.000	-.149
	Sig. (2-tailed)	.491	.108	.821	.	.058
	N	100	69	162	163	163
Corrective Maintenance (Purified)	Pearson Correlation	.102	.455**	.048	-.149	1.000
	Sig. (2-tailed)	.127	.000	.361	.058	.
	N	223	155	360	163	378

**. Correlation is significant at the 0.01 level (2-tailed).

117

Regression Results for Corrective Maintenance

Model Summary[b,c]

Model	R		R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
	INCLUDE2 = 1.00 (Selected)	INCLUDE2 ~= 1.00 (Unselected)				R Square Change	F Change	df1	df2	Sig. F Change
1	.799[a]	.712	.639	.615	.4458	.639	26.515	3	45	.000

a. Predictors: (Constant), Tech Assist Visit Requests per System, Sailor Proofness (averaged), Natural Logarithm of MTBF

b. Unless noted otherwise, statistics are based only on cases for which INCLUDE2 = 1.00.

c. Dependent Variable: Maintenance Workload (Purified)

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15.808	3	5.269	26.515	.000[a]
	Residual	8.943	45	.199		
	Total	24.750	48			

a. Predictors: (Constant), Tech Assist Visit Requests per System, Sailor Proofness (averaged), Natural Logarithm of MTBF

b. Dependent Variable: Maintenance Workload (Purified)

c. Selecting only cases for which INCLUDE2 = 1.00

118

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	2.855	.539		5.293	.000	1.769	3.941		
	Natural Logarithm of MTBF	-.320	.079	-.395	-4.042	.000	-.479	-.160	.842	1.187
	Sailor Proofness (averaged)	-.338	.111	-.290	-3.034	.004	-.562	-.113	.877	1.140
	Tech Assist Visit Requests per System	1.175	.231	.467	5.090	.000	.710	1.640	.954	1.048

a. Dependent Variable: Maintenance Workload (Purified)

b. Selecting only cases for which INCLUDE2 = 1.00

Collinearity Diagnostics^{a,b}

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	Natural Logarithm of MTBF	Sailor Proofness (averaged)	Tech Assist Visit Requests per System
1	1	3.524	1.000	.00	.00	.00	.02
	2	.435	2.846	.00	.00	.01	.89
	3	3.384E-02	10.205	.08	.06	.97	.01
	4	7.382E-03	21.849	.92	.94	.01	.07

a. Dependent Variable: Maintenance Workload (Purified)

b. Selecting only cases for which INCLUDE2 = 1.00

119

Residuals Statistics[a,b]

| | INCLUDE2 = 1.00 (Selected) | | | | | INCLUDE2 ~= 1.00 (Unselected) | | | | |
	Minimum	Maximum	Mean	Std. Deviation	N	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-1.0960	1.5759	.1579	.5739	49	-.9455	1.0530	1.573E-02	.5429	12
Residual	-.9648	.9666	-4.53E-17	.4316	49	-.6742	.6521	-3.56E-02	.4516	12
Std. Predicted Value	-2.185	2.471	.000	1.000	49	-1.923	1.560	-.248	.946	12
Std. Residual	-2.164	2.168	.000	.968	49	-1.512	1.463	-.080	1.013	12

a. Dependent Variable: Maintenance Workload (Purified)

b. Pooled Cases

Charts

Histogram of Residuals

Regression Standardized Residual

Std. Dev = .97
Mean = 0.00
N = 49.00

120
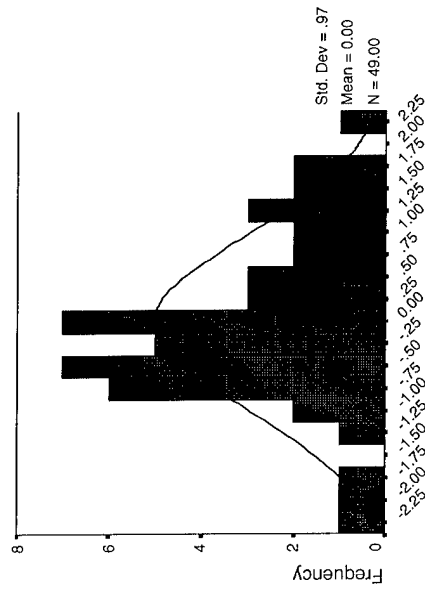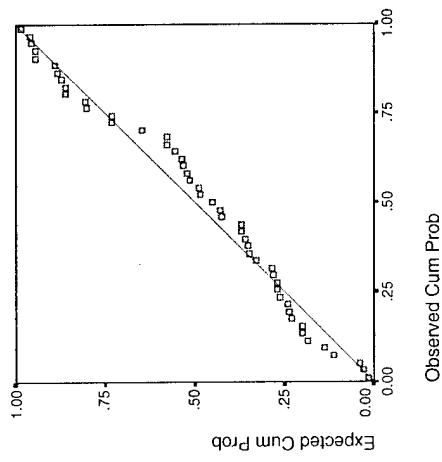
Note "Maintenance Workload (Purified)" denotes Corrective Maintenance in the plots below.

Normal P-P Plot of Standardized Residuals

Partial Regression Plot

Partial Regression Plot

Maintenance Workload (Purified)

Sailor Proofness (averaged)

□ Unselected Cases

Selected Cases

Partial Regression Plot

Maintenance Workload (Purified)

Tech Assist Visit Requests per System

□ Unselected Cases

Selected Cases

122